

Marthe Aune  
Julie Røed

## Master Thesis

BI Norwegian Business School

# **- Forced Ranking: Friend or Foe? -**

On forced ranking and its effect on intrinsic motivation,  
justice perceptions and performance

Hand-in date:  
01.09.2011

Campus:  
BI Oslo

Examination code and name:  
GRA 19002 Master Thesis

Program:  
Master of Science in Leadership and Organizational Psychology

Supervisor:  
Thorvald Hærem

This thesis is part of the MSc program at BI Norwegian Business School. The school takes no responsibility for the methods used, results found and conclusions drawn.

---

**Table of Content**

<b>Acknowledgements</b> .....	<b>iii</b>
<b>Abstract</b> .....	<b>iv</b>
<b>1. Introduction</b> .....	<b>1</b>
<b>2. Theory and Hypotheses</b> .....	<b>4</b>
<b>2.1 Forced Ranking</b> .....	<b>4</b>
2.1.1. Forced Ranking and Feedback .....	4
2.1.2. Relationship Patterns of Forced Ranking and Performance .....	5
2.1.3. Feedback and Performance .....	6
<b>2.2. Intrinsic Motivation</b> .....	<b>8</b>
<b>2.3. Justice Perceptions</b> .....	<b>10</b>
<b>2.4. Conceptual Model</b> .....	<b>12</b>
<b>3. Method</b> .....	<b>14</b>
<b>3.1. Experimental Task</b> .....	<b>14</b>
<b>3.2. Procedure</b> .....	<b>14</b>
<b>3.3. Sample</b> .....	<b>15</b>
<b>3.4. Operationalizations</b> .....	<b>15</b>
3.4.1. Forced Ranking.....	15
3.4.2. Performance .....	15
3.4.3. Intrinsic Motivation, Distributive- and Procedural Justice .....	19
3.4.6. Factor analysis.....	19
3.4.7. Control variables .....	21
<b>4. Analysis</b> .....	<b>23</b>
<b>4.1. Assumptions of HLM</b> .....	<b>24</b>
<b>4.2. Multicollinearity</b> .....	<b>25</b>
<b>4.3. Analytic strategy</b> .....	<b>25</b>
<b>5. Results</b> .....	<b>27</b>
<b>5.1. Descriptive Statistics</b> .....	<b>27</b>
<b>5.2. Hypothesis Testing</b> .....	<b>28</b>
<b>5.2. Control Variables</b> .....	<b>30</b>

---

<b>5.3. Mediation Testing .....</b>	<b>30</b>
<b>5.4. Interaction Effects .....</b>	<b>31</b>
<b>6. Discussion .....</b>	<b>34</b>
<b>6.1. Forced ranking.....</b>	<b>34</b>
<b>6.2. Intrinsic Motivation .....</b>	<b>35</b>
<b>6.3. Justice Perceptions .....</b>	<b>36</b>
<b>6.4. Control Variables.....</b>	<b>38</b>
<b>7. Limitations.....</b>	<b>40</b>
<b>8. Implications .....</b>	<b>42</b>
<b>8.1. Implications for practice .....</b>	<b>42</b>
<b>8.2. Implications for future research .....</b>	<b>43</b>
<b>9. Conclusion .....</b>	<b>44</b>
<b>Reference list.....</b>	<b>45</b>
<b>Appendices .....</b>	<b>50</b>
<b>APPENDIX A – Measures .....</b>	<b>50</b>
<b>APPENDIX B – Control Variables .....</b>	<b>52</b>
<b>APPENDIX C – Questionnaire, Control variables .....</b>	<b>54</b>
<b>APPENDIX D – The tasks .....</b>	<b>56</b>

***Acknowledgements***

First and foremost we would like to thank our supervisor Thorvald Hærem. Through the whole process of writing this master thesis he has shown genuine interest in our project and us. He has offered theoretical guidance, practical support and statistical competence, but maybe more importantly he has unfailingly believed in us. We are truly grateful for having him as our supervisor.

Bjørn Bakken has also been invaluable in this process. This project would not have managed without him. We would really like to thank him for his support and advises and for all the time and effort he has put into this project.

We also want to thank Anders Dysvik and Bård Kuvaas for their theoretical insights and advices. Their seemingly endless tank of knowledge on motivation and feedback has been of great help.

Finally, writing a master thesis entails some hard prioritizing and long hours. We would therefore really like to thank the people closest to us for their support and love. A special thank to Bård Egil for his great ability to listen and how he somehow always manages to make dark clouds go away. And to Ine, my one and only, you are the sunshine of my life.

Marthe Aune

Julie Røed

**Abstract**

This study explores the effects of forced ranking, conceptualized as a summative, norm-referenced form of feedback, on performance. It was hypothesized that the lower ranked individuals would demonstrate less performance improvement than the higher ranked individuals after receiving feedback. The results were opposite as hypothesized, with higher ranked individuals performing worse compared to lower ranked individuals. These results are explained and discussed in light of both earlier and more recent feedback theories. To further investigate the effects of forced ranking, three mediating variables were included. The first mediating variable, intrinsic motivation, had only a marginally significant effect on performance. Forced ranking also failed to reach statistical significance on intrinsic motivation, and mediation is therefore not supported. The second and third mediating variables were distributive and procedural justice. Forced ranking did not achieve a significant effect on these variables. The effect of distributive- and procedural justice on performance did not produce significant results either, which rejects the mediation hypotheses. Nonetheless, post hoc interaction analyses discovered a significant interaction effect of procedural justice on the relationship between forced ranking and performance; participants with lower perceptions of procedural justice performed better than those with higher perceptions, and this effect was the most pronounced when higher ranked. A possible explanation for this rather unconventional finding is discussed, together with explanations and discussions of the insignificant and significant findings of all variables.

In sum, some of the propositions of feedback theory are supported by this study, and the facilitating tendency of intrinsic motivation on performance is to some degree further established. Nevertheless, in order to sort out of the effects of forced ranking on performance and other work-related variables, more research is clearly needed.

## 1. Introduction

Among the human resource practices, performance appraisals is regarded as one of the most important (Judge & Ferris 1993, Boswell & Boudreau 2002), and one of the most frequently used in organizations today (Blume, Baldwin & Rubin 2009). Performance appraisals include, according to Boswell and Boudreau (2002), tasks such as identification of an individual's strengths and weaknesses, goal setting, and recognition of training needs. Typically, these evaluations support other human resource activities such as promotion and pay administration, and are characterized by comparing an individual's performance to a standard, other members of the organization or previous performance. However, despite their pervasive use, research has identified many problems in relation to the application of performance appraisals. Rating errors stemming from rater bias has in particular received attention. Rater bias is the tendency on part of the raters to give lenient or inflated ratings. Consequently, this bias leads to a lack of differentiation between high and low performers (Blume, Baldwin & Rubin 2009). Given such inflation, it is argued that performance evaluations lose their credibility, as they fail to differentiate between the employees (Guralnik, Rozmarin & So 2004). Recently, it has therefore been a revival of forced ranking systems (Blume, Baldwin & Rubin 2009), which is a type of performance appraisal that separate individuals into preexisting performance categories (Olson & Davis 2003; Hazels & Sasse 2008).

Being about categorizing people, forced ranking has been and still is a controversial issue. This has resulted in a heated debate over the pros and cons in both professional HR journals and the media (e.g. Meisler 2003). Jack Welch, the former superstar CEO of General Electric, is one of the most famous proponents of forced ranking (or the "vitality curve" as referred to in the Welchian lingo), arguing that forced ranking is the key to the organization's competitive advantage (Schleicher, Bull & Green 2009). Nonetheless, respected authors have questioned the validity and effectiveness of the forced ranking approach (Pfeffer & Sutton 2006). The debate however, is more founded on anecdotal accounts (e.g. Lawler III 2002; Grote 2005), than on empirical

academic research (Blume, Baldwin & Rubin 2009). As a response to this debate several authors have pointed to the need for research on forced ranking (Scullen, Bergey & Aiman-Smith 2005; Blume, Baldwin & Rubin 2009). There are several unanswered questions regarding forced ranking, the most prominent one is perhaps whether this practice actually increases the performance of employees. A key issue is therefore to identify how forced ranking influence performance. This study examines how receiving different rank affects performance, and therefore contributes to the human resource literature in general, and to the performance appraisal literature in particular.

Given that there is a finding between forced ranking and performance it is of interest to include possible mediating variables in order to understand and better explain the potential finding. This as mediators speak to why and how effects between independent and dependent variables occur (Baron & Kenny 1986). In this study three mediating variables are introduced to broaden our understanding of the relationship between forced ranking and performance.

First, forced ranking can be regarded as a type of feedback as feedback can be defined as a form of communication that conveys some degree of information about past behavior, performance or achieved understanding (Hattie & Timperley 2007; Ilgen, Fisher & Taylor 1979). As feedback is widely believed to affect motivation (Bandura 1993; Deci & Ryan 2000; Locke & Latham 2002), and subsequent performance (Callahan, Scully, Brownlee, Brtek & Tosi 2003) it is reason to believe that the relationship between forced ranking and performance is mediated by motivation. The link between motivation and feedback is evident in many motivational theories (Schunk, Pintrich & Meece 2008). In self-determination theory this relation is very much apparent. This theory proposes the concept of intrinsic motivation, which occurs when a person performs an activity merely for itself (Deci & Ryan 1985). Intrinsic motivation is related to feedback in that feedback can spur or diminish the feeling of autonomy and competence, and thus either increase or decrease a person's intrinsic motivation

(Gagné & Deci 2005). Therefore, we propose intrinsic motivation to mediate the relationship between forced ranking and performance.

Second, in the debate over forced ranking, justice perceptions have been identified as a potentially important variable (Lawler III 2002; Meisler 2003; Olson & Davis 2003; Sears & McDermott 2003). This is because forced ranking makes the ranking of an individual's performance dependent on others' performance. Roch, Sternburgh and Caputo (2007) suggest that such relative performance appraisal formats are perceived to be less fair than absolute formats. As such, it is reason to believe that justice perceptions, more specifically distributive and procedural justice, mediate the relationship between forced ranking and performance. On this basis we propose the following research question:

*How does forced ranking affect intrinsic motivation, justice perceptions and individual performance?*

The paper adheres to the following outline. First, the next section presents the theoretical background and the hypotheses to be tested in order to examine our research question. A conceptual model is introduced to illustrate the hypothesized relationships. The method section outlines the procedure of the study, operationalizations and measurement of variables and method of analysis. The results of the hypotheses testing is then presented and then subsequently discussed in relation to the theoretical background. We also performed a post hoc interaction analysis, which is presented in the same section. Then we present our reflections on limitations of this study and possible implications. Finally, the main findings are summarized in a conclusion.



## **2. Theory and Hypotheses**

In this section we review research and theories on forced ranking, feedback, intrinsic motivation and justice perceptions. On this basis we suggest hypotheses.

### ***2.1 Forced Ranking***

Forced ranking is a type of performance appraisal where evaluations are required to fit along the lines of a particular distribution (Schleicher, Bull and Green 2009). This performance evaluation approach is based on the repeated finding in social sciences that when measured in large enough samples, most human phenomena tend to follow a normally distributed curve (Guralnik, Rozmarin & So 2004). The “archetype” of forced ranking is thus the procedure of categorizing individuals into preexisting performance categories, against other employees in the department or peer group (e.g. a 20-70-10 distribution) (Olson & Davis 2003). These performance rankings are then applied to a bell curve, with those ranking at the bottom (usually 10%) being put on probation, given improvement possibilities or terminated. By contrast, those ranking on top (usually 20%) are generously rewarded for their performance (Hazels & Sasse 2008). Obviously, forced ranking systems contrasts with an absolute system of evaluation, in which employees are evaluated on the basis of an absolute standard and not in relation to other ratees (Duffy & Webber 1974).

#### ***2.1.1. Forced Ranking and Feedback***

Clearly, performance ratings such as forced ranking provide an important source of feedback to individuals in organizations (Bartol, Durham & Poon 2001). Despite impassioned anecdotal accounts (e.g. Lawler III 2002; Grote 2005) on both side of the debate, very little empirical research has emerged on forced ranking (Blume, Baldwin & Rubin 2009). Therefore, it is of interest to look into what the feedback literature could contribute with to this debate. On a general level, feedback is a form of communication that conveys some degree of information about past behavior, performance or achieved understanding (Ilgen,

Fisher & Taylor 1979; Hattie & Timperley 2007). Yet having this stated it is important to note that feedback is far from a simple stimulus as feedback has several dimensions. One of the most important distinctions is feedback sign-whether the feedback is positive or negative (Ilgen, Fisher & Taylor 1979; Podsakoff & Farh 1989). Furthermore, feedback can be conceptualized as formative or summative and norm-referenced or self-referenced (Chan & Lam 2010). Summative feedback focuses on the outcome, whereas formative feedback provides the individual with learning cues in how to progress (Taras 2005; Covic & Jones 2008). Self-referenced feedback involves self-comparative appraisal, whereas norm-referenced concerns social-comparative appraisal (Chan & Lam 2010). Applied to the case of forced ranking it seems reasonable to argue that forced ranking is a summative, norm-referenced form of feedback that is distributed according to a variant of the normal distribution (e.g. 20-70-10). The feedback sign will depend on what ranking the individual receives.

### ***2.1.2. Relationship Patterns of Forced Ranking and Performance***

The little attention that has been devoted to the field of forced ranking also transmits to our knowledge about the relation between forced ranking and performance. Nonetheless, there are some studies that have approached the issue of forced ranking and performance. A simulation study conducted by Scullen, Bergey and Aiman-Smith (2005) investigated if implementation of a forced distribution rating system (FDRS) could improve the average quality of an organization's workforce. Their findings revealed that FDRS could in fact improve the workforce potential, however, potential side effects such as decline in employee moral, general dissatisfaction, lowered organizational commitment and possible increase in turnover were identified. Research conducted by Garcia and Tor (2007) provides further knowledge on why these negative effects might occur. By nature, forced ranking involves social comparison. Leaning on research and findings from Festinger, Garcia and Tor (2007) claim that this comparison process often results in competitive behavior. Their findings indicate that it is competition on a general scale rather than task comparison that is the main social comparison facilitator of competitive behavior, and therefore suggest that

forced ranking can lead to greater competition among the employees. Greater competition may sound positive; however this can actually have a detrimental effect. As stated by Garcia and Tor (2007, 106):

while highly ranked employees may be more competitive and productive through simple self selection, the championing of forced rankings fails to anticipate how competitive forces may ultimately inhibit the profit-maximizing exchange or pooling of information and resources among those 'star' employees.

On the basis of the study by Scullen, Bergey and Aiman-Smith (2005) it is possible to argue that there is a positive relationship between forced ranking and organizational performance. However, we are interested in how forced ranking relates to individual performance, which makes the findings of this study less relevant. As forced ranking is conceptualized as a type of feedback, we turn back to the feedback literature to investigate the findings concerning feedback and performance.

### ***2.1.3. Feedback and Performance***

It is a well-established finding that feedback is related to performance (Kim & Hamner 1976; Illgen, Fisher & Taylor 1979; Larson Jr. 1989; Early, Northcraft, Lee & Lituchy 1990; Kluger & DeNisi 1998; Goodman, Wood & Hendrickx 2004; Hattie & Timperley 2007; Anseel, Lievens & Schollaert 2009). As pointed to above, forced ranking is a form of performance feedback, hence forced ranking are expected to relate to performance in some way or another. The initial theoretical arguments for the effectiveness of feedback were provided by Thorndike and his law of effect (Kluger & DeNisi 1998). Positive feedback was equated with reinforcement, and negative feedback with punishment. Both types of feedback should improve performance because positive feedback reinforces performance, whereas negative feedback punishes the erroneous behavior (Kluger & DeNisi 1998). An influential and much cited review by Ammons (1956) gave further support for the beneficial effect of feedback on performance. Given this understanding of feedback we would expect that feedback provided through a forced ranking system would be beneficial to performance because those ranked

high should be even more motivated to perform, whereas those ranked as average- or low performers would get a kick in the pants to enhance their performance.

Nevertheless, more recent research on feedback (e.g. Kluger and DeNisi 1996) suggests that not all feedback necessarily lead to better performance. Indeed, the presence of negative effects of feedback is robust; about 34-38 % of the effect sizes investigated in a meta-analysis by Kluger and DeNisi (1996) showed a negative effect on subsequent performance. The theoretical explanation provided (Feedback Intervention Theory) suggests that feedback that directs attention to the self (for example "You are a great student") is more likely to attenuate the effect of feedback on performance. By contrast, feedback effects on performance are augmented by feedback that is related to the task (for example "This essay can be improved if elaborating more on the theoretical concepts"). The explaining mechanism is that cues that shifts attention to the self reallocates cognitive resources from the task to the self, and in such a way weaken performance (Kluger & DeNisi 1996). The major discriminator is thus whether feedback is directed to the task or to the self level (Hattie & Timperley 2007). Moreover, grading research also supports that feedback in form of grades could have a negative effect on performance. Although Cherry and Ellis (2005) found that rank-order grading could generate improved student performance relative to criterion- referenced grading, Butler and Nisan (1986) found that grades might encourage an emphasis on quantitative aspects of learning, reduce creativity, promote fear of failure, and weaken interest. As we observe, the picture gets more complicated because the effect of feedback is not as straightforward as was hypothesized in earlier theory and research.

The distinction concerning whether feedback is directed to self or task is useful in relation to forced ranking. Given that feedback derived from forced ranking is norm-referenced, that is, feedback that conveys comparative information, it could be argued that this type of feedback diverts attention from the task to the self. Feedback that directs attention to the self via normative cues has been

shown to be largely ineffective (Kluger and DeNisi 1998). Similarly, a study by Butler (1987) found that grades increased ego involvement, but did not affect performance relative to the no-feedback control group. This contradicts the earlier understanding of feedback, in that feedback is not universally positively linked to performance (Kluger & DeNisi 1998). Nonetheless, it could also very well be argued that forced ranking feedback also conveys information regarding task performance, although on a relative scale. In contrast to feedback directed to the self, task- focused feedback has shown to increase task involvement and consequently performance (Butler 1987). In particular, feedback that provides corrective information (e.g. formative feedback) has shown to be effective in relation to performance (Hattie & Timperley 2007). Thus, we run into muddy waters, as the forced ranking feedback does not neatly fall into the distinction between feedback directed to task or self. Consequently, it is somewhat difficult to predict the effects of forced ranking on performance. However, given that forced ranking conveys comparative information, it would be hard to overlook the social comparison aspect of forced ranking, which is argued to be largely ineffective in improving performance (Kluger & DeNisi 1996). Furthermore, as negative feedback is found to be more potent than positive feedback at the self level (Hattie & Timperley 2007), we argue that participants receiving an average (middle) or low ranking will decrease their performance after receiving feedback. Based on the above account, we therefore hypothesize the following:

***H1. The low and middle ranked individuals will demonstrate less performance improvement than highly ranked individuals after receiving feedback.***

## **2.2. Intrinsic Motivation**

Owing to the limited research on forced ranking, we know little about what factors that might intervene between forced ranking and performance. However, as forced ranking represents a type of feedback, we expect that the effect of forced ranking on performance could be mediated by motivation. This as feedback is widely believed to affect motivation (Bandura 1993; Deci & Ryan 2000; Locke & Latham 2002), which subsequently affects performance by

directing attention and increasing persistence and effort (Callahan, Scully, Brownlee, Brtek & Tosi 2003).

Intrinsic motivation is a core tenant in Deci and Ryan's self-determination theory (SDT) and cognitive evaluation theory (CET). CET was presented by Deci and Ryan as a sub-theory within SDT with the aim of specifying factors that explain variability in intrinsic motivation (Deci & Ryan 2000). Intrinsic motivation can be defined as the motivation to perform an activity for itself, in order to experience the pleasure and satisfaction inherent in the activity (Deci & Ryan 1985). According to the STD and CET, feedback that are interpreted as information about one's competence and satisfy individuals' need for autonomy will enhance intrinsic motivation (Deci, Ryan & Koestner 1999; Gagné & Deci 2005).

Moreover, research has shown that positive performance feedback can enhance intrinsic motivation, and that negative performance feedback can diminish it (Deci, Ryan & Koestner 1999; Deci & Ryan 2000). However, positive feedback that is perceived as controlling, that is, positive feedback having an evaluative character, or emphasizing how one should perform – has clearly shown to decrease intrinsic motivation (Ryan 1982; Deci, Connell & Ryan 1989). Ryan (1982) suggests that positive feedback can be perceived either as informational or controlling depending on various factors, and that these will determine whether the positive feedback increases or decreases intrinsic motivation. Drawing this link to forced ranking it is reasonable to argue that positive ranking could be interpreted as controlling. This as forced ranking has a normative character and as such state something about how a person should perform, which may lead to a decrease in intrinsic motivation. However, a positive rank could also be interpreted as information about one's competence and therefore increase intrinsic motivation. For participants' receiving a negative rating it is possible to infer that they will experience a decline in intrinsic motivation in both of the conditions – either they perceive the ranking as controlling or as decreasing perceptions of competence: both of which decrease intrinsic motivation. Accordingly, there is reason to believe that the low and average

(middle) ranked individuals will experience lower levels of intrinsic motivation than the higher ranked.

Before the millennium few studies had examined the performance effects associated with intrinsic motivation. Rather, much of the intrinsic motivation literature had focused on how extrinsic motivational sources affect intrinsic motivation (Callahan, Scully, Brownlee, Brtek & Tosi 2003). Today however we have achieved a great deal more knowledge on this relation. For example, Callahan, Scully, Brownlee, Brtek and Tosi (2003) examined the unique effects of multiple sources on task performance, and found that intrinsic motivation had the greatest effect on performance. Moreover, two studies by Kuvaas (2006; 2007) reported a positive relation between intrinsic motivation and performance. In a study on transformational leadership and job behaviors Piccolo and Colquitt (2006) also found that the indirect effect of intrinsic motivation supported the direct effect of transformational leadership on task performance. Finally, a study by Dysvik and Kuvaas (2008) observed that the relationship between perceived training opportunities and work performance was fully mediated by intrinsic motivation. All these studies thus suggest that intrinsic motivation is a potent predictor of task performance. Accordingly we hypothesize the following:

***H2.** The relationship between forced ranking and performance will be mediated by intrinsic motivation.*

### **2.3. Justice Perceptions**

Justice has been identified as a potentially important variable in the debate over forced ranking (Lawler III 2002; Meisler 2003; Olson & Davis 2003; Sears & McDermott 2003), and a peak into the organizational justice research therefore seems appropriate. Research on organizational justice has identified different forms of justice, most notably distributive justice and procedural justice (Colquitt 2001). Distributive justice refers to the perceived fairness of outcomes, whereas procedural justice refers to the perceived fairness of the processes by which

outcomes where arrived at (Cohen-Charash & Spector 2001). A vast literature provides evidence for people making distinctive judgments about procedural and distributive justice, and that both types of judgments can predict behaviors, decisions, or evaluations in important ways (Lucas 2009).

Concerning the relation between forced ranking and justice perceptions, there is little empirical research. Although considerable research has documented the importance of justice perceptions in connection to performance evaluation processes in organizations (Bartol, Durham & Poon 2001), there is however no studies to our knowledge that investigates justice perceptions and subsequent performance after receiving forced ranking feedback. Nonetheless, research on performance evaluations has identified the format of performance appraisal systems to be important in connection to justice. A study by Roch, Sternburgh and Caputo (2007) suggest that relative formats are perceived to be less fair than absolute formats, with the forced ranking format perceived to be the least fair. Furthermore, Bartol, Durham and Poon (2001) link justice with the segmentation of performance appraisal systems. They point to that a typical three-category system is designed to capture 70-80 % of employees in the middle category. For example, if performance were normally distributed, an employee receiving performance feedback one standard deviation below average would typically receive the same performance rating as a colleague performing one standard deviation above average, which could influence justice perceptions. Thus, it does not seem unreasonable to argue that forced ranking could influence justice perceptions.

It is sensible to argue that both procedural and distributive justice could be affected by forced ranking. First, research findings suggest that people tend to be highly influenced by social comparison information, and that information about an individual's standing within a group influences distributive justice perceptions (Bartol, Durham & Pool 2001). Forced ranking does indeed convey information regarding relative standing within a group; hence it is argued that forced ranking affects the perceptions of distributive justice. Given the tendency for people to



rate themselves above average (Meyer 1975), it is argued that people receiving a low- or average (middle) rank will perceive distributive justice to be lower than the highly ranked. Second, research has shown that when outcomes are low, perceptions of procedural justice becomes more important (Roch, Sternburgh & Caputo 2007). It is proposed that individuals who receive high ratings may not be particularly concerned about procedural justice, and therefore would be more likely to perceive the ranking procedure as fair. The low and middle ranked will on the other hand be more concerned about procedural justice, and therefore perceive less procedural justice than the higher ranked.

Regarding the relation between justice perceptions and performance, a number of studies have linked justice perceptions to important organizational outcomes such as affective commitment (Kuvaas 2003), organizational commitment (Farndale, Hope-Hailey & Kelliher 2011), performance (Lind, Kanfer & Earley 1990) and turnover (Simons & Roberson 2003). Furthermore, equity theory holds that when an individual perceives distributive injustice at work, the employee can alter his or her quality and quantity of work to re-establish justice (Cohen-Charash & Spector 2001). Thus, it would be in the best interest of organizations to maximize employees' justice perceptions (Roch, Sternburgh & Caputo 2007). Based on the above accounts we therefore hypothesize the following:

**H3.** *Procedural justice will mediate the relationship between forced ranking and performance*

**H4.** *Distributive justice will mediate the relationship between forced ranking and performance*

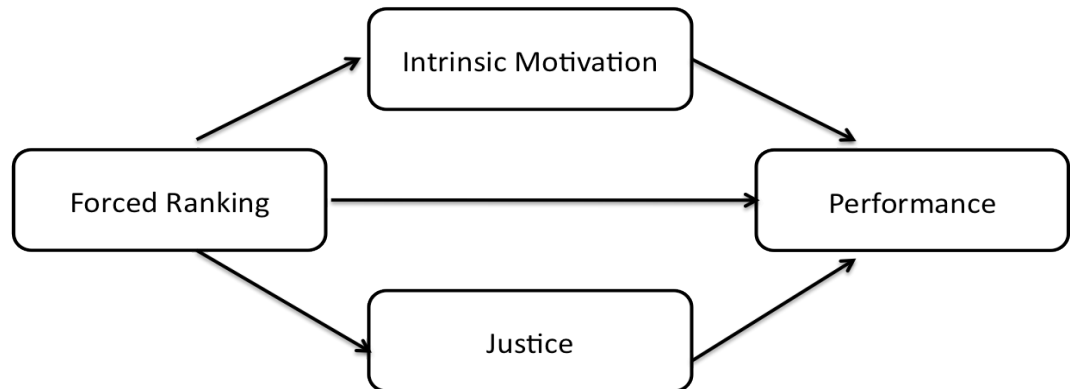
#### **2.4. Conceptual Model**

The conceptual model (Figure 1) is based on the preceding presentation and discussion of the hypotheses. H1 is based on the general feedback literature and propose that lower ranked participants will demonstrate less performance improvement than the higher ranked after receiving feedback. H2 are based on

intrinsic motivation as a mediating variable between forced ranking and performance. H3 and H4 are based on procedural- and distributive justice as mediating the relationship between forced ranking and performance.

**Figure 1**  
*Conceptual Model*

---



### **3. Method**

#### ***3.1. Experimental Task***

The present study was based on a computer simulation allowing for participants to become ranked in correspondence to their performance relative to others. Computer simulations moreover provide participants with a complex model of reality (Salas, Wildman, and Piccolo 2009), and are therefore considered suitable for the purpose of this study as this complexity might trigger intrinsic motivation, as an intrinsically motivated person seeks out novelty and challenges (Deci & Ryan 2000).

In the simulation the participant's task was to monitor a map of south of Norway, where they had to handle incidents that 'popped up' as blinking signs on the map. By clicking on these signs the participants received text messages describing each incident and cues on how to proceed. In order for the participants to handle the incidents the participants were instructed to make use of several resources that they could engage by 'dragging' them from their current locations on the map, and 'drop' the resources on the blinking sign (incident symbol).

There were four types of resources available in the computer simulation: transportation helicopters, rescue helicopters, surveillance aircraft and fighter aircraft. For each resource it was possible to select a particular capacity that could be more suitable for each of the incidents. A default capacity for each resource was automatically assigned if no capacity was selected. The resource would start to move towards the incident symbol (the blinking sign) as soon as a resource was engaged to an incident. In total there were twelve incidents in each of the scenarios and both lasted for approximately twenty minutes.

#### ***3.2. Procedure***

Upon arrival all of the participants received instructions about the simulation and were told that the experiment had something to do with behavior within

organizational psychology. Then they were given a test scenario to practice all of the functions in the game, before completing two scenarios (which was the 'real' experiment). Both prior to, and during the experiment the participants received questionnaires that they had to complete.

### **3.3. Sample**

Participants in this study were mainly from BI Norwegian Business School and consisted of 80 participants. In total, six sessions were run with 7 to 18 individuals participating each time. Participants in this study were between 21 and 41 years of age, and 76.3% of the participants were female. 77.5% had a Norwegian nationality and 78.8% had Norwegian as their first language. 31.2% had some military experience.

### **3.4. Operationalizations**

#### **3.4.1. Forced Ranking**

After completing scenario 1 all participants were given feedback on their performance relative to the other participants by receiving information on whether their performance was low, middle or high. The participants were in other words distinguished into three levels. The levels were based on a 25-50-25 distribution, which was communicated to all of the participants. This distribution was chosen for practical reasons as this distribution required fewer participants than a 10-70-20 distribution.

#### **3.4.2. Performance**

The performance scores in each scenario were determined by whether the participant managed to react quickly (decision speed), and with accuracy (selecting the right resources and capacities). At the end of each task in the scenario, the participant's degree of success at handling the situation was displayed as a numeric "effect" variable ranging from 0 to 100%. If the task were successfully solved, the incident symbol would turn green, or turn black if the

effect was lower than 75 %. In total, we have twenty-four observations of performance per participant, as each completed twelve tasks both in scenario one and in scenario two.

Unfortunately, there are missing values on our performance variables. When presence of missing values, the issue is to identify the patterns underlying the missing data in order to maintain as close as possible the original distribution of values when any sort of remedy is applied (Hair, Black, Babin & Anderson 2010). Two questions are of particular interest: 1) Are the missing data randomly distributed or distributed in distinct patterns? 2) How prevalent are the missing data? Concerning the first question, our missing data is a result of technological difficulties (game breakdown) and as such not tied to either the independent or the dependent variable. Therefore, the distribution of missing data has to be characterized as MCAR (missing completely at random) (Tabachnick & Fidell 2007). This is supported by a non-significant Little's MCAR test ( $\chi^2 = 57.55$ ,  $df = 57$ ,  $Sig. = 0.455$ ). The null hypothesis for this test is that the data are missing completely at random, and a statistical non-significant result is therefore desired.

Turning to the prevalence of the missing data, our missing data analysis (MVA) reveals that we have a total of 145 missing cases out of 1920 observations ( $24 \times 80 = 1920$ ). The missing data make up 7,55 % of the total data. However, the missing values are clustered on 7 of the 24 performance variables, making the percentage of missing values much higher for some of the variables (see Table 1).

**Table 1**  
*Missing values per variable*

Variable	No. of cases	Percentage %
PERF1S11	18	22,50 %
PERF1S12	18	22,50 %
PERF2S8	1	1,25 %
PERF2S9	18	22,50 %
PERF2S10	18	22,50 %
PERF2S11	36	45 %
PERF2S12	36	45 %

*Note.* PERF1Sx = Scenario 1, PERF2Sx = Scenario 2.

According to Tabachnick and Fidell (2007), it is an option to drop variables with missing values when the missing values are concentrated in few variables, contingent that these are not critical to the analysis. PERF2S11 and PERF2S12 have such a high number of missing values that we find it reasonable to exclude them from further analysis. Excluding these two implies deletion of PERFIS11 and PERF1S12 as well, because we need comparable observations. This leaves us with only three variables with missing values: PERF2S8 (1 missing), PERF2S9 (18 missing) and PERF2S10 (18 missing). Given that we think it would be unreasonable to delete these variables because of the resulting data loss, we are left with either deleting the missing cases or impute the missing data (Hair, Black, Babin & Anderson 2010). As our missing values are characterized as MCAR we have many options.

The listwise method uses only cases with complete data (Hair, Black, Babin & Anderson 2010). This approach is however limited in use for our data, because it would reduce the sample size by nearly one quarter, which is a massive loss of data. Therefore we need to consider imputation as a means to replace the missing values. Imputation is the process of substituting the missing values based on valid values of other variables and/or cases in the sample (Hair, Black, Babin & Anderson 2010). Our choice of imputation method fell on regression imputation, because it is a more sophisticated method for estimating missing values (Tabachnick & Fidell 2007). Other variables are here used as independent variables to estimate a regression equation for the variables with missing values serving as the dependent variables. It is not without its disadvantages; this method understates variance unless an error term is added to the replacement values (Tabachnick & Fidell 2007). Fortunately, SPSS (the statistical software used) allows adding a random component to the regression estimates. The regression imputation was therefore run with adding residuals to the regression estimates. An inspection of the frequencies of the three variables before and after imputation was thereafter performed, in order to verify that the regression method had produced sensible values.

With the imputation procedure successfully performed, the most appropriate way to calculate the performance variable had to be decided on. Given that we are not interested in the absolute performance level of participants but rather the change in performance from time one to time two, some sort of change variable had to be created. We chose not put performance at time two as the dependent variable, and performance at time one as a covariate in addition to the other independent variables. The reason why is that performance time one and one of the independent variables, forced ranking, contain to a very large degree the same information, and therefore could potentially inflate the effect of forced ranking.

A second concern was adjusting for regression to the mean. Regression to the mean (RTM) refers to "...the tendency for extreme observations in a distribution at baseline to move closer to the mean at follow-up." (Smith & Beaton 2008, 290). Applied to our case, it is not unlikely that individuals performing excellent at time one could perform less well in the second scenario, and vice versa for the individuals performing very bad, regardless of the performance feedback (ranking) actually received. Thus, if RTM is not adjusted for, we run the risk of not comprehending the effects of forced ranking feedback because RTM could potentially distort our analysis. Therefore, the change in performance from time one (scenario 1) and time two (scenario 2) were measured as residual scores, as this method has the advantage of not inflating error that might arise with the use of difference scores (Schaufeli, Bakker & Rhenen 2009). The residual scores from the regression indicate who has improved more, or less, than expected based on their initial baseline score of performance (Smith & Beaton 2008). Following the recommendations by Smith and Beaton (2008) the residuals scores were calculated by regressing time two scores of performance on the corresponding time one scores, after centering the performance at time one variable. In addition, a squared term of performance at time one was added to the regression equation to better adjust the residuals relative to each participant. However, as modeling with the squared term did not produce substantially

different results from modeling without, the final regression equation of performance did not include the squared term.

### ***3.4.3. Intrinsic Motivation, Distributive- and Procedural Justice***

Intrinsic motivation, distributive and procedural justice were measured by a questionnaire before scenario 2, right after the participants had received feedback. The intrinsic motivation scale used in the present study was based on a scale developed by Kuvaas and Dysvik (2009) and was adapted to fit the context of the simulation. Participants were asked to indicate their agreement with six items concerning their inner drive to accomplish their tasks on a five point Likert scale. The distributive and procedural justice scales used in the present study were based on two scales developed by Colquitt (2001) and were adapted to fit the context of the simulation. Participants were asked to indicate their agreement with a total of eleven items concerning their feedback score and the procedures to arrive at that feedback score on a five point Likert scale. The questionnaire with items measuring intrinsic motivation, distributive- and procedural justice are presented in Appendix A.

### ***3.4.6. Factor analysis***

The items of the intrinsic motivation, distributive justice and procedural justice scales (17 in total) were subjected to principal components analysis with oblique rotation (Direct Oblimin). The sample size of 80 was just below the recommendations of at least five participants per variable ( $5 \times 17 = 85$ ) (Field 2009). However, the value of the Kaiser-Meyer- Olkin Measure of Sampling Adequacy (KMO) was 0.784, which is well above the recommended value of 0.6 (Pallant 2010), and all KMO values for individual items were  $> 0.6$ , which is above the acceptable limit of 0.5 (Field 2009). The Bartlett's Test of Sphericity reached statistical significance ( $p = .000$ ). These measures suggest that our data set is appropriate for factor analysis (Pallant 2010). An initial analysis was run to obtain eigenvalues for each component in the data. Four factors had eigenvalues above Kaiser's criterion of 1, and in combination explained 69.8 % of the



variance. The scree plot was slightly ambiguous, and showed inflexions that would justify retaining both three and four components. Given that we on theoretical grounds expected three underlying factors to emerge, this is the number of factors retained in the final analysis.

**Table 2**

*Pattern matrix and Communalities of Three Factor Solution of Intrinsic Motivation, Distributive Justice and Procedural Justice<sup>a</sup>*

Item	Pattern coefficients			Communalities
	1	2	3	
The tasks that I did in the simulation were themselves representing a driving power	.756			.590
The tasks that I did in the simulation was enjoyable	.855			.773
I felt that the simulation was meaningful	.857			.739
The simulation was very exciting	.851			.828
The simulation was so interesting that it was a motivation in itself	.849			.771
I was so inspired by the simulation that I almost forgot everything around me	.823			.638
Does your feedback score reflect the effort you have put into the simulation activity?		.862		.708
Is your feedback score appropriate for the activity you have completed?		.790		.648
Does your feedback score reflect what you have contributed with in the simulation activity?		.867		.771
Is your feedback score justified, given your performance?		.695		.641
Do you think that those procedures have been applied consistently?			-.692	.551
Do you think those procedures has been free of bias?			-.889	.726
Do you think those procedures have been based on accurate information?			-.770	.653
Do you think that those procedures have upheld ethical and moral standards?			-.624	.521

Extraction Method: Principal Component Analysis.

a. Rotation converged in 5 iterations.

We rerun the analysis with specifying three as the number of components to extract. The three items (Procedural justice items 1, 2 and 6) that previously made up the forth factor performed unsatisfactorily in this solution, and were

subsequently dropped because they failed to reach the level of statistical significance recommended for our sample size: 0.6 (Hair, Black, Babin & Anderson 2010). These items were not critical to our analysis. Finally, the analysis was performed without the deleted items. The three retained components explained in combination 68.2 % of the variance, with Component 1 explaining 35.4 %, Component 2 explaining 20.5% and Component 3 contributing 12.4 %. The rotated solution revealed the presence of simple structure (Pallant 2010), with all components showing a number of strong loadings and all variables loading considerably on only one component. The interpretation of the three components is straightforward; intrinsic motivation items loaded strongly on Component 1, distributive justice items on Component 2, and procedural justice items on Component 3. The pattern matrix and communalities of the final solution is presented in Table 2. There were weak correlations between the factors (see Table 3). Note that these correlations are based on an Oblimin rotation, and that the interpretation of the relation between these variables is only appropriate within the factor analysis. The variables to be used in the later analyses are generated on equally weighted items from the raw data, and the correlations will therefore be different (see table 4). The results of this analysis support the use of the items as separate scales. For the factor intrinsic motivation, a Cronbach's alpha value of .919 is well above the recommended value of .7 (Hair, Black, Babin & Anderson 2010). The Cronbach's alpha for distributive justice and procedural justice scales were .825, and .766 respectively.

**Table 3**  
*Component Correlation Matrix<sup>a</sup>*

	Intrinsic motivation	Distributive justice	Procedural justice
Intrinsic motivation	1	-	-
Distributive justice	.043	1	-
Procedural justice	-.252	-.219	1

<sup>a</sup>Extraction method: Principal Component Analysis

### **3.4.7. Control variables**

Uncontrolled extraneous variables may pose a serious threat to validity, hence it is important to measure them in order to control for their effect on the

dependent variable (Pedhazur & Schmelkin 1991). Control variables were measured by a questionnaire sent out via email to participants beforehand (see Appendix C). The control variables that were measured by several items were subjected to factor analysis, and internal consistency of the items were estimated in order to make sure that a single construct was measured. See Appendix B for these procedures.

**Strategic computer game experience.** Experience with strategic computer games can make a difference in a simulation that involves tasks that requires strategic thinking. Therefore, this was included as a control variable.

**Military experience.** In line with the argument above, the military setting of the computer simulation may benefit those with military experience. Military experience was measured with four items. These were subjected to factor analysis, and the results revealed the presence of two factors: the first reflecting crisis experience and the other military education. The Cronbach's alpha of the first factor was .667. The other factor consisted of only one item, thus internal consistency reliability cannot be estimated.

**Demographic variables.** Age, gender, first language, and nationality were measured to estimate and examine their effect on the dependent variable.

**Task difficulty.** The tasks in each scenario vary in difficulty. In both scenarios, the first three tasks are characterized as easy, along with task number five, six and seven. The remaining tasks (no. 4, 8, 9, 10, 11 and 12) were considered difficult. As we decided to delete task 11 and 12 because of their high levels of missing values, this leaves us with four tasks that are difficult and six tasks that are considered easy. Task difficulty was included in order to control for the level of demand associated with each of the tasks.

## 4. Analysis

Hierarchical linear modeling (HLM) was used for analysis in the present study. HLM was chosen for two reasons. First, because we have a repeated measures data set, and secondly because the data is organized on more than one level. When data for participants is organized on more than one level, the individual observations are generally not independent as the observations from the same individual are commonly more similar than observations from another individual (Hox 2010). Figure 2 illustrates the levels of measurement and the measurements performed at each level. The first level is the repeated measures level, with measurements of the ten tasks in each scenario (N=800). The second level constitutes the individual participants (N=80). The data are nested as the repeated observations at the first level are clustered in individuals at the second level.

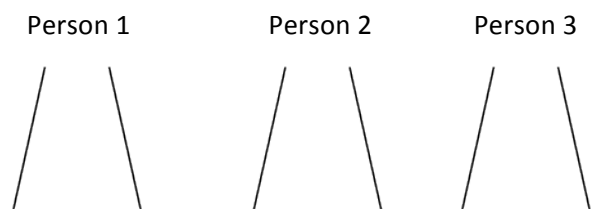
**Figure 2**

*Levels of Measurement and Measurements Performed at Each Level*

---

Level 2: Persons (N = 80)

Predictor: Ranking, Intrinsic  
Motivation, and Justice



Level 1: Performance (N = 800)

Task 1 . . . . Task 10    Task 1 . . . . Task 10    Task 1 . . . . Task 10

In repeated measures data the dependent variable is measured more than once for each participant, and HLM is often used for repeated measurements with the repeated measurement at the first level of analysis (Hox 2010). This is because multiple observations are available for each participant, and these observations tend to be correlated with each other (West 2009). By organizing the data on different levels we thus avoid violating the assumption of independence of all observations and interpretational errors as well. For example, organizing the data on only one level would most likely in our case imply aggregation of the

dependent data, which potentially could lead to loss of information (Stevens 2009). HLM relaxes the assumption of independence by allowing for predictors at every level of analysis, and for means and relationships between higher units to vary. HLM thus permits both fixed and random effects between variables. For random effects the value of a parameter is different for each subject whilst for fixed effects the value of a parameter is 'fixed' to a constant over all subjects (Tabachnick & Fidell 2007). Random effects are therefore specific to subjects within the population, whilst fixed effects describe the relationships between the dependent variable and predictor variables for an entire population (West, Welch & Galecki 2007). In the present study random effects were estimated for the repeated measures of performance (level 1) and persons (level 2). We consider persons to be a random factor as they were randomly sampled from a larger population. The remaining variables (forced ranking, intrinsic motivation, distributive and procedural justice) were estimated as fixed effects.

#### ***4.1. Assumptions of HLM***

According to Tabachnick and Fidell (2007), a sample size of 60 is required in HLM when about five parameters are estimated. Given that we have 7-8 parameters to estimate in each model, our sample size of 80 with ten observations on performance per person, giving a total of 800 observations on performance, is deemed sufficient. As described in section 3.4 the missing values were replaced with imputation. Performing the analysis without the imputed data did not produce noticeable differences, which indicates that the imputation was appropriate. Furthermore, as HLM is an extension of multiple linear regression, the assumptions of this technique also pertains to HLM (Tabachnick & Fidell 2007). In order to assess normality, a descriptive statistic was run on all the predictors and the dependent variable. All variables met the assumptions of normality, linearity and homoscedasticity, except for intrinsic motivation that displayed signs of kurtosis. Nonetheless, an inspection of the histogram and normal probability plot showed that this variable was reasonably normal (Pallant 2010). Transformations were therefore not performed. Inspection of boxplots

detected no outliers, and all cases had acceptable Mahalanobis distance values (Tabachnick & Fidell 2007).

#### **4.2. Multicollinearity**

Correlated predictors are problematic in HLM because the effects of correlated predictors are adjusted for, which makes it more likely that none of their regression coefficients will be statistically significant (Tabachnick & Fidell 2007). A multicollinearity statistic was therefore run on the variables. Of the predictors, the lowest tolerance value was .74, which is far above the recommended threshold value of .10 (Hair, Black, Babin & Anderson 2010). Of the control variables, the lowest tolerance value was .248, which also is above the recommended threshold value of .10. Finally, all predictors except rank (only three levels) were centered. Centering was performed to avoid issues with multicollinearity when performing the post hoc interaction analysis (Tabachnick & Fidell 2007).

#### **4.3. Analytic strategy**

The hypothesis testing was performed with HLM in two steps. First, a null model for the dependent variable was estimated. Second, the full model with all predictors and the significant control variables was estimated. Each hypothesis is then examined through evaluating the sign, size and significance level of each regression parameter. Overall model fit is evaluated by comparing the values of the Akaike Information Criterion (AIC) of the two models. AIC is used because this information criterion penalizes the log likelihood for number of parameters estimated (Singer 1998). In order to evaluate which of the models that have the best fit, the following procedure will be adhered to: The minimum AIC of the two models will be identified and then rescaled with a value of 0. As such, we obtain an estimate of the size of the information loss for the other model compared to the best model of the two (Burnham & Anderson 2002). This estimate will then be evaluated in light of the recommendations of Burnham and Anderson (2002), which suggest that models having an AIC difference within 1-2 of the best model

have substantial support, models within 4-7 of the best model have less support, and finally, models with an AIC difference of  $> 10$  have minimal support and as such fail to explain some substantial structural variation in the data. Finally, the method of estimation was maximum likelihood (ML) for both models, as the hypothesis testing involves comparison of the two models (Tabachnick & Fidell 2007).

## 5. Results

### 5.1. Descriptive Statistics

Table 4 presents the descriptive statistics and the intercorrelations among the independent and dependent variables. Note that since the correlations are based on the HLM dataset all the variables have a sample of 800 in order to include all the observations on performance. To clarify, this means that performance have 800 different observations, whilst the other variables have 80 x 10 observations. This makes these variables more likely to become significant as they have ten times higher sample size than they really have. This is because large samples are overly sensitive to significance testing, detecting almost any relationship as statistically significant (Hair, Black, Babin & Anderson 2010).

**Table 4**

*Means, Standard Deviations, and Intercorrelations of the Independent and Dependent Variables*

Variable	M	SD	1	2	3	4	5
1. Forced ranking	1.02	.725	—				
2. Performance	.000	.291	-.098**	—			
3. Intrinsic Motivation	.000	5.562	.140**	.041	—		
4. Procedural Justice	.000	3.144	.011	-.032	.348**	—	
5. Distributive Justice	.000	3.730	-.131**	-.015	.053	.262**	—

\*  $p < .05$ . \*\*  $p < .01$ .

Table 4 indicates no support for H1, as forced ranking is negatively correlated with performance ( $r = -.098$ ,  $p < .01$ ), which is opposite of the hypothesized direction. There is some preliminary support for hypothesis H2, as there is a significant positive correlation between forced ranking and intrinsic motivation ( $r = .140$ ,  $p < .01$ ), however, there is no significant relationship between intrinsic motivation and performance. Hypothesis H3 receives no support, given the non-significant correlation between forced ranking and procedural justice, and procedural justice and performance. Finally, hypothesis 4 also receives no preliminary support, as there is a significant negative correlation between forced ranking and distributive justice ( $r = -.131$ ,  $p < .01$ ), which is opposite of what was hypothesized, and a non-significant correlation between distributive justice and performance.



## 5.2. Hypothesis Testing

Table 5 presents the fixed regression parameters, their standard errors and statistical significance. Table 7 presents the AIC values for each model. The AIC difference is > 10, which indicate that the predictors as a group did improve the model beyond the null model.

**Table 5**  
*Multilevel Modeling with Dependent Variable and Control Variables*

Variable	Performance	Intrinsic mot.	Procedural jus.	Distributive jus.
Forced ranking	-.037** (.011)	n/a <sup>a</sup>	n/a <sup>a</sup>	n/a <sup>a</sup>
Intrinsic mot.	.003 <sup>1</sup> (.001)			
Procedural jus.	-.004 (.003)			
Distributive jus.	.001 (.002)			
Task difficulty	.344** (.014)			

*Note.* The regression parameters appear above the standard errors (in parentheses).

<sup>a</sup> No estimate of regression parameter because model failed to converge.

<sup>b</sup> Task difficulty is a categorical variable coded as dummy variable with two levels; difficult = 0, easy = 1.

\*  $p < .05$ . \*\*  $p < .01$ .

<sup>1</sup>  $p < .10$

**Table 6**  
*Simple regressions between Forced ranking and Mediators*

Variables	Intrinsic mot.	Procedural jus.	Distributive jus.
Forced ranking	.140	.011	-.131

*Note.* The regression coefficients are standardized coefficients.

\*  $p < .05$ . \*\*  $p < .01$ .

**Table 7**  
*Comparison of Multilevel Models for Performance*

Model	AIC	$\Delta$ AIC
1. Null model	- 8.988	+ 77, 094
2. Full model	- 86.082	0

From Table 5 it is obvious that the effect of forced ranking, intrinsic motivation, procedural justice and distributive justice on performance is mixed. On average, the change in performance is negatively related to forced ranking level ( $p < .01$ ). This implies that those ranked high performed worse after receiving forced ranking feedback, and that those ranking low performed better after receiving forced ranking feedback. It was hypothesized that lower ranked participants would demonstrate less improvement than higher ranked participants, Hypothesis 1 therefore receives no support.

The multilevel model failed to converge when estimating the effect of forced ranking on intrinsic motivation. However, as can be seen in Table 6, a simple regression was run to estimate this relationship. Although the sign and size of the regression coefficient indicates a positive relationship between forced ranking and intrinsic motivation, the regression coefficient is not significant, indicating no statistical evidence for the theoretical assertion that forced ranking affects intrinsic motivation. Furthermore, it was hypothesized that individuals with high levels of intrinsic motivation would perform better than individuals with low levels. The effect of intrinsic motivation on performance is marginally significant, which indicates a positive relation between intrinsic motivation and performance ( $p = .065$ ). Nevertheless, as our significance level a priori was set to .05, we have no indisputable statistical evidence for this relation.

Turning to the justice variables, the multilevel model also failed to converge when estimating the effects of forced ranking on procedural and distributive justice. Simple regressions were therefore run (see Table 6), with the result of non-significant regression coefficients for both procedural and distributive justice. The regression coefficient for procedural justice is quite small, but still

positive, indicating the tendency that forced ranking affect procedural justice in the hypothesized direction. Nonetheless, the regression coefficient is non-significant, providing no statistical evidence for the relationship between forced ranking and procedural justice. The regression coefficient for forced ranking on distributive justice is larger than that for procedural justice, and unexpectedly, negative in sign. This is an unexpected finding, and clearly not in the hypothesized direction. Regarding the justice variables on performance, procedural and distributive justice did not significantly affect performance ( $p = .187$ ,  $p = .627$ , respectively). In addition, the regression coefficient for procedural justice was in the opposite direction, indicating a negative relationship between procedural justice and performance. Taken together, these findings imply rejection of Hypothesis 3 and Hypothesis 4.

### **5.2. Control Variables**

Table 5 also points to significant effects of control variables. Of all control variables, task difficulty was the only one with statistical significance ( $p < .01$ ). This is a categorical variable with difficult tasks in the scenario coded as 0 and easy tasks as 1. The regression parameter of .344 suggests that performance is higher when the tasks are easy, which was expected.

### **5.3. Mediation Testing**

Based on the framework of Baron and Kenny (1986), a variable functions as a mediator when 1) variations in levels of the independent variable significantly account for variations on the hypothesized mediator, 2) variations in the mediator significantly explain variations in the dependent variable, 3) a previously significant relationship between the independent and dependent variable becomes insignificant when the mediator is controlled for. First, forced ranking must affect the mediators (intrinsic motivation, procedural- and distributive justice). When we tried to model these relationships the model failed to converge, leaving us with no reliable estimate of these relationships. The simple regressions produced non-significant results, indicating that forced ranking does not have an effect on neither intrinsic motivation, nor procedural or

distributive justice. Second, the mediators must also affect the dependent variable in order to establish mediation. Here there is only a marginally significant relationship between intrinsic motivation and performance, and not significant relationships between the justice variables and performance. Third, it is vital that there is a relationship between the independent variable forced ranking and the dependent variable performance. There is a significant relationship here, although not in the hypothesized direction. Given that the only significant relationship on a .05 level is between forced ranking and performance, the conditions are not present to establish mediation. Thus, none of the mediation hypotheses (2, 3 and 4) are supported. Due to this state of affairs of non-significance we will conduct post hoc tests to investigate if there are any interaction effects. Given that such moderating effects are found, they could have high informational value in terms of specifying what individual differences that should be taken into consideration when forced ranking systems are implemented and/or evaluated.

#### ***5.4. Interaction Effects***

In general terms, a moderator is a qualitative or quantitative variable that affects the direction and/or strength of the relationship between a predictor and a dependent variable (Baron & Kenny 1986). In order to test for a moderator effect, three paths must be estimated: 1) the relation between the predictor and dependent variable, 2) the impact of the moderator on the dependent variable, and 3) the interaction of these two. The moderator hypothesis is supported if the interaction term is significant (Baron & Kenny 1986).

In relation to intrinsic motivation, it is reasonable to argue that it could be conceptualized as a moderator of the relationship between forced ranking and performance. Although intrinsic motivation is influenced by situational factors such as task characteristics, it could be argued that it has a global component as well (Kuvaas 2006). Thus, it is possible that some individuals are dispositionally more likely to be intrinsically motivated than others. Such individuals may be less

affected by forced ranking, because they in general engage in an activity for itself and are not too concerned with external factors such as forced ranking.

Furthermore, it could also be argued that both procedural and distributive justice could have a moderating effect on the relationship between forced ranking and performance. With basis in the Just World Theory, it is suggested that beliefs about procedural and distributive justice not only encompass situational assessments of justice, but also more stable dispositional tendencies to perceive outcomes and/or rules and processes as uniquely deserved (Lucas 2009). More specifically, individuals with high distributive justice beliefs are more likely to believe that people generally get what they deserve in life. Similarly, individuals with high procedural justice beliefs are more likely to perceive that people in general are treated fairly (Lucas 2009). Given that participants' answering on procedural and distributive justice items in our experiment reflect such deeper attitudes, it could be argued that participants with high distributive justice beliefs would be more likely to perceive forced ranking as fair because it could be argued to be a reflection of what people deserves. Accordingly, individuals with high beliefs in procedural justice are perhaps also more likely to perceive forced ranking as rather fair, because they believe that they and people in general are treated fairly.

**Table 8**  
*Interactions among Independent variables on Performance*

Variable	Intrinsic mot.	Procedural jus.	Distributive jus.	Full model
Forced ranking	-.004 (.002)	-.007* (.003)	-.004 (.002)	
AIC	- 86.733	- 87.454	- 85.933	-86.082
$\Delta$ AIC	+ 0.721	0	+ 1.521	+ 1.372

*Note.* All models estimated with all predictors and the interaction term entered separately.

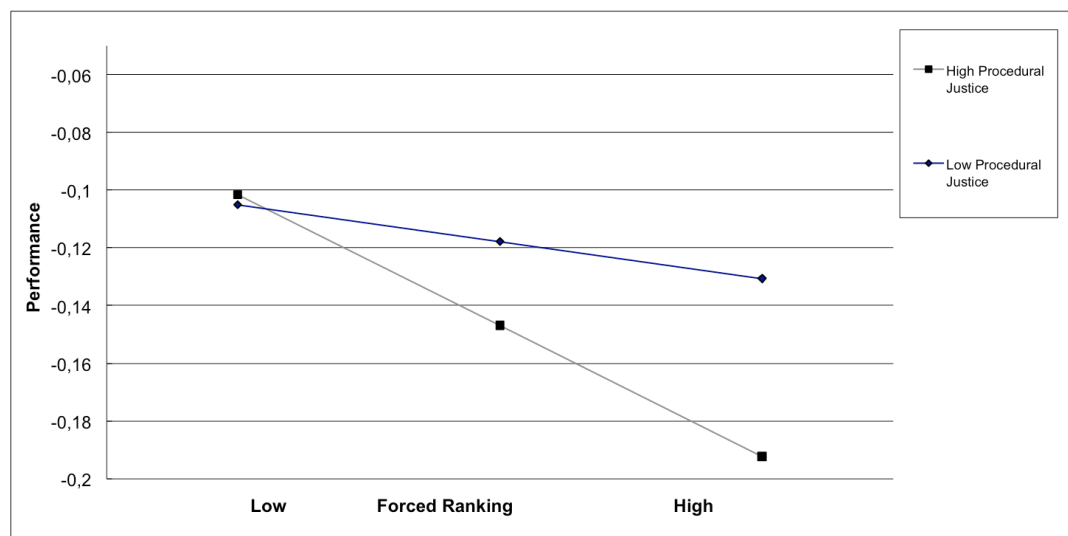
\*  $p < .05$ .

In order to evaluate how the added interactions affected model fit, the model with the minimum AIC is rescaled to 0, with the purpose of establishing a baseline comparison model. As can be observed from attending to Table 8, the interaction term between forced ranking and procedural justice is the one with

the lowest AIC value, and the only one reaching statistical significance ( $p = .044$ ). For graphical illustration of the interaction see Figure 3. Given the non-significance of the other two interaction terms, these moderator hypotheses receive no support (although the AIC values are acceptable).

**Figure 3**

*Interaction between forced ranking and procedural justice on performance*



*Note.* Performance is conceptualized as the expected change in performance from time one to time two and not in absolute terms (see discussion page 15-19).

As can be seen from Figure 3, two rather puzzling findings are observed. First, confirming the negative sign of the regression coefficient for procedural justice on performance (see Table 5), those with higher perceptions of procedural justice perform worse in the second scenario than those with lower perceptions of procedural justice. Given our theoretical expectations that perceiving high procedural justice would be beneficial for performance, this finding clearly speaks in the opposite direction. Second, this tendency is increasing with rank level, suggesting that this effect is most pronounced when individuals are ranked high. Thus, when ranked high, those with higher perceptions of procedural justice perform worse in scenario two after receiving forced ranking feedback than those with low procedural justice.

## 6. Discussion

In this study, we have investigated the relationships between forced ranking, intrinsic motivation, justice perceptions and performance. Several hypotheses were proposed, and in the following these will be discussed in light of the findings just presented.

### **6.1. Forced ranking**

Based on the tenets of feedback intervention theory, it was hypothesized that the lower ranked individuals would demonstrate less performance improvement than the higher ranked individuals after receiving feedback. This was grounded in the argument that forced ranking is a summative, norm-referenced form of feedback, which is more likely to direct attention to the self-level, which in turn has been found to be largely ineffective (Kluger & DeNisi 1998). Individuals that were ranked low were expected to show less performance improvement because these participants received feedback with a negative sign. The results however, tell a different story. Contrary to our predictions, individuals ranking high performed worse after receiving feedback, whereas those ranking low performed better than those ranking high in scenario two.

Interestingly, this finding can be explained by combining the two theoretical frameworks of earlier feedback theory, and the more recent one of feedback intervention theory. Thorndike and his law of effect suggest that both positive and negative feedback should improve performance because positive feedback reinforces performance, whereas negative feedback punishes the erroneous behavior (Kluger & DeNisi 1998). By contrast, feedback intervention theory (Kluger & DeNisi 1996) holds that normative feedback that directs attention to the self could attenuate performance. Our findings indicate that receiving positive feedback through being ranked high does not bring forward any facilitating effect on subsequent performance, thus not supporting the principles of earlier feedback theory. This finding is more in line with feedback intervention theory, which suggest that feedback that provides comparative cues that shifts

attention to the self is not so effective. Our findings suggest that such feedback is not only ineffective; it may also lead to deterioration in performance for those highly ranked. Such a finding is interesting because it suggests that being ranked high is not a “bed of roses”, at least regarding subsequent performance.

On the other hand, receiving negative feedback through being ranked low might not be so destructive as suggested by some authors (e.g. Hattie and Timperly 2007). In fact, those who were ranked low improved performance versus those who were ranked high in the next scenario, and this clearly does not support our hypothesis that lower ranked individuals would demonstrate less performance improvement than the higher ranked individuals. This finding might indicate that these individuals actually did get a kick in their pants to improve their performance because of the low ranking, supporting the tenets of earlier feedback theory. In sum, even though the results are in the complete opposite direction of what was hypothesized, this study still supports the well-established finding of feedback research: performance feedback does seem to have an effect on subsequent performance.

### **6.2. Intrinsic Motivation**

Based on general feedback literature, intrinsic motivation was suggested to mediate the relationship between forced ranking and performance. The results showed a non-significant relationship between forced ranking and intrinsic motivation, and a marginally significant positive relationship between intrinsic motivation and performance. The finding that intrinsic motivation is positively related to performance is far from novel. Several studies, amongst them Dysvik and Kuvaas (2008), have identified intrinsic motivation as a potent predictor of task performance. Although the finding between intrinsic motivation and performance was only marginally significant, this finding still adds to this line of research and provides further support for, and credit to, the notion that intrinsic motivation enhances performance. Regarding the relationship between forced ranking and intrinsic motivation, the simple regression found that this relationship was positive, and as such indicates the same direction as



hypothesized. However, as it is not significant we have no statistical evidence to establish that there really is such a relationship. Consequently, mediation is not supported either.

One reason for why the relationship between forced ranking and intrinsic motivation is not significant might be subscribed to the manipulation of forced ranking. The forced ranking was not tied to any consequences and might therefore not have been perceived as very controlling. It is perhaps more likely that being ranked would have been perceived as more controlling if there had been any consequences tied to the different rankings, as the normative character of forced ranking might have felt more outspoken then. Another reason could be the characteristics of forced ranking. The summative aspect of forced ranking does not provide much information and might therefore not have influenced feelings of competence in either direction. Given that either of the reasons is the case, or both, they might explain why we cannot establish intrinsic motivation as a mediator.

### ***6.3. Justice Perceptions***

In the debate over forced ranking, justice has been identified as a potentially important variable (Lawler III 2002; Meisler 2003), and it was therefore hypothesized that both procedural and distributive justice would mediate the relationship between forced ranking and performance. According to our results, this however does not seem to be the case. Regarding the relationship between forced ranking and distributive justice, the obtained regression coefficient between forced ranking and distributive justice was in the opposite direction as hypothesized, and not of significance. Turning to procedural justice, we find that this relationship (forced ranking – procedural justice) is far from being significant, however, the regression coefficient is in the hypothesized direction. Thus our findings indicate that forced ranking do not trigger justice perceptions in this study. One explanation for this finding might be subscribed to the time span of the experiment. According to Cohen-Charash & Spector (2001), Leventhal and colleagues state that six criteria should be met in order for a procedure to be

perceived as fair. One of these criteria states that procedures should be applied consistently across people and across time. As our experiment only lasted for a few hours it might be that the time period was too short for procedural justice to kick in. The same explanation might also transfer to distributive justice in accordance with the vast amount of research that has found high correlations between procedural and distributive justice (Colquitt, Conlon, Wessen, Porter & Ng 2001). Another possible explanation for our insignificant findings might also here be related to the lack of consequences tied to the forced ranking. According to Colquitt, Conlon, Wessen, Porter and Ng (2001), Adams suggested that one way to determine whether an outcome is fair is to calculate one's own 'input' to one's 'output' and then compare that ratio with the perceived ratio of others. The 'output' in our experiment is to get a rank that is labeled as high-, middle- or low. Receiving one of these rankings does however not imply any consequences. It might be that this lead to a too weak output and that the participants therefore did not go through with the calculation, which might have lead to non-significant findings.

Regarding the relationship between the justice variables and performance, it was hypothesized that higher perceptions of both distributive and procedural justice would have a beneficial effect on performance. Both justice variables failed to reach statistical significance. In addition, the relationship between procedural justice and performance was opposite of the hypothesized direction, indicating a negative relationship between procedural justice and performance. In explaining these findings, the results of a meta-analysis by Cohen-Charash and Spector (2001) may come in handy, as they found a striking difference between laboratory and field studies outcomes when dealing with work performance related to justice. The authors found that whereas the results of field studies showed a strong relationship between procedural justice and performance, results of laboratory studies showed a weak relationship between the two and no relationship between distributive justice and performance. Cohen-Charash and Spector suggest an explanation for this finding that might relate to our study as well, providing a possible explanation for our non-significant finding. They

suggest that in a laboratory setting, the relationship between justice and performance is much weaker than in field studies because performance is influenced by situational demands that are salient in the laboratory, more than in the field. In our experiment, the participants were instructed to make use of several resources when solving missions, and that error in solving missions would be marked by a black sign and correct ones with green. These instructions indicated what to do in order to solve the tasks successfully, and the black or green sign denoted actual success in accomplishing the missions. In this respect it might be that the performance requirements in our laboratory setting is more salient than in a field setting, which might lead to more feelings of injustice in a field setting because it might be more confusion around what it takes to perform well there.

Finally, the significant interaction found between forced ranking and procedural justice on performance did not shed any further light on why there was found a negative relationship between procedural justice and performance. It was found that those with higher perceptions of procedural justice perform worse after receiving forced ranking feedback than those with low procedural justice, and that this effect is strongest when highly ranked. It is challenging to provide an explanation for this finding, as it is difficult to find convincing arguments for why participants who perceive that they have been treated fairly and have received a high rank should perform worse in the second scenario than the rest of the participants. Nevertheless, one possible explanation might be that those who experienced procedural injustice performed better because they were more aroused by the feelings of injustice, and consequently engaged more in the tasks in the second scenario. This is of course pure speculation, however it could be an interesting issue to explore for future research.

#### **6.4. Control Variables**

Strategic computer game experience, military experience, demographic variables and task difficulty were included as control variables. Of these, only task

difficulty was found to be significant. The easier tasks were easier to solve and the more difficult tasks were more difficult to solve. This finding was expected.

## 7. Limitations

As with all research, this study has some limitations. First, the manipulation of forced ranking might have been too weak. Organizations vary in terms of the purpose for conducting forced ranking, however the ranking is normally tied to some kind of consequence. At one extreme, organizations use forced ranking to terminate the lowest performing employees'. At the other end of the scale, some organizations collect the ratings for record keeping purposes only. The more common use is for organizations to use forced ranking to determine promotions or demotions, different assignments and compensation (Schleicher, Bull & Green 2009). In our experiment we did not have the opportunity to tie any consequences to the different rankings, which may have caused a too weak manipulation. Furthermore, forced ranking typically also follow a 20-70-10 distribution (Olson & Davis 2003; Hazels & Sasse 2008), whereas we chose a 25-50-25 distribution for practical reasons. This might also have affected the strength of the manipulation. As such, it might be that some of the hypothesized effect of forced ranking disappears when there is no consequences tied to the given rank and when utilizing a less strict distribution.

A second limitation of this study is related to the strength of the manipulation check. After the participants received their rank we asked them if they were aware of the rank they had received. However, this manipulation check could have been stronger in that we for example had asked them to write down their rank. This would have given us a greater certainty that they were aware of the rank they received. As such, we recommend future research to consider a stronger manipulation check for forced ranking.

A third limitation is subscribed to the external validity of this study. As previously discussed, the manipulation of the forced ranking might have been too weak in that the setting of the experiment differs from a more natural setting. This may restrict the generalizability of the findings from the present study. However, this limitation is to some extent leveled out by the internal validity of this study. A great strength of this study is that we know when the feedback was given,

namely after the participants had conducted scenario one and that this potentially affected the performance in scenario two. Thus, speaking in causality terms, we have at least established the independent variable taking place before the dependent variable, which makes our inferences about the effect of forced ranking feedback more credible.

Finally, we might not have been capable of removing regression to the mean completely from our analysis. Given that the lowly ranked performed better and that the highly ranked performed worse after feedback, it is reasonable to argue that this also could be due to regression to the mean. As explained in the method section, attempts were made to adjust for regression to the mean, however, a complete removal of RTM is probably not achieved. As such, we cannot be absolutely sure that the observed tendency for highly ranked to perform worse in the second scenario, and for the lower ranked to perform better is in fact only due to the forced ranking performance feedback. Indeed, RTM would perhaps be a better explanation of the tendency for the highly ranked individuals to perform worse after feedback, as there is no strong theoretical basis for claiming that positive feedback should lead to the observed deterioration in subsequent performance.

## 8. Implications

### ***8.1. Implications for practice***

The implications for practice are especially associated with the importance of forced ranking in improving performance. Our results indicate that higher ranked individuals perform worse after receiving such feedback. This implies that it may not be so unproblematic to be ranked highly as suggested on theoretical grounds. According to Kluger and DeNisi (1998), this type of feedback should be largely ineffective, however in our case this type of feedback seems to lead to deterioration in performance. Coupled with the tendency (although non-significant) that highly ranked individuals perceive less distributive justice than individuals with lower rank, it might indicate that highly ranked individuals may not feel comfortable being ranked high. This may to some degree be supported by a study by Blume, Baldwin and Rubin (2009), in which four elements of forced ranking system was investigated, and a key finding was that respondents were most attracted to systems with less stringent treatment of low performers. This might suggest that people do have a concern for the low performers, even though they are ranked high. This is of course speculation, however, it might be worth being aware of and take into consideration if implementing or conducting a forced ranking system.

Another practical implication is related to our findings on intrinsic motivation. Our results found a positive marginally significant relationship between intrinsic motivation and performance, however, not a significant relationship between forced ranking and intrinsic motivation. Although the relationship between intrinsic motivation and performance is only marginally significant, it still has some practical value. This finding implies that investment in peoples' intrinsic motivation might be beneficial for performance. Thus, this finding might be relevant for leaders and people who are working within human resources who want to stimulate a high performing work environment. For a social environment to maximize intrinsic motivation it should provide people with the opportunity to satisfy their basic psychological needs for competence, relatedness and

autonomy (Deci, Vallerand, Pelletier & Ryan 1991). It is therefore important that performance appraisals emphasize feedback that satisfies the individuals' need for competence and autonomy (Gagné & Deci 2005).

### ***8.2. Implications for future research***

Our findings imply that forced ranking is negatively related to performance. This suggests that there might be several downsides of being ranked high and perhaps not so bad to be ranked low, which to some degree can be explained by more recent feedback theory (e.g. Kluger & DeNisi 1996) and earlier feedback theory. However, these findings unfortunately do not add much clarification to our understanding of forced ranking, and still there are many unanswered questions for future research to address. In particular, it would be interesting to see a laboratory experiment utilizing a strict distribution (e.g. 20-70-10), and that tied real consequences to the ratings. For example, if one or several grades of the students participating in this study were directly linked to what ranking they received, we might have observed much stronger effects of forced ranking.

Furthermore, the findings from this study did not support the inclusion of mediators. However, it would probably be shortsighted to conclude that there is no need for future research to include mediators in order to understand the effects of forced ranking on performance. The non-significance of the mediators may be due to the manipulation being too weak, and an area of future research would therefore be to investigate both intrinsic motivation and justice perceptions in conditions in which the outcome of forced ranking is tied to real consequences for the employees.



## 9. Conclusion

As a confrontation to the lenient performance appraisal systems, the practice of forced ranking emerged. The increasing popularity of forced ranking calls for knowledge on its effects. Drawing on previous and more recent feedback theory, this study has examined the influence of forced ranking on performance and aimed to enhance our understanding of this relationship by introducing three mediating variables. We found a significant relationship between forced ranking and performance, however, the relationship was opposite of our prediction. Although this finding was in the opposite direction as hypothesized, it still adds to feedback theory and research by indicating that performance feedback such as forced ranking seems to have an effect on performance, which is a well-established finding in feedback research.

The introduction of the mediating variables intrinsic motivation, procedural- and distributive justice revealed no further clarification as almost all of these relationships were found to be non-significant. An exception is the marginally significant finding of intrinsic motivation on performance, which gives some preliminary support to a substantial body of research suggesting a facilitating effect of intrinsic motivation on performance. Post hoc interaction analyses also discovered a significant interaction effect of procedural justice on the relationship between forced ranking and performance. This finding, together with the rest of the non-significant findings, leaves us more questions than answers. Chiefly, we wonder why forced ranking failed to have any effect on our mediators, in addition to why lower ranked individuals performed better relative to higher ranked individuals in scenario two.

In sum, some of the propositions of feedback theory are supported by this study, and the facilitating tendency of intrinsic motivation on performance is to some degree further established. Nevertheless, in order to sort out the effects of forced ranking on performance and other work-related variables, more research is clearly needed.

---

## Reference list

- Ammons, Robert B. 1956. Effects of knowledge of performance: A survey and tentative theoretical formulation. *Journal of General Psychology*, 54: 279-299.
- Ansel, Frederik, Filip Lievens and Eveline Schollaert. 2009. Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes*, 110: 23-35
- Bandura, Albert. 1993. Perceived Self- Efficacy in Cognitive Development and Functioning. *Educational Psychologist*, 28(2): 117- 148.
- Baron, M. Reuben and David A. Kenny. 1986. The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51(6): 1173-1182.
- Bartol, Kathryn, Cathy C. Durham and June M. L. Poon. 2001. Influence of Performance Evaluation Rating Segmentation on Motivation and Fairness Perceptions. *Journal of Applied Psychology*, 86(6): 1106-1119.
- Blume, Brian D., Timothy T. Baldwin and Robert S. Rubin. 2009. Reactions to Different Types of Forced Distribution Performance Evaluation Systems. *Journal of Business and Psychology*, (24): 77-91.
- Boswell, Wendy R. and John W. Bourdreau. 2002. Separating the Developmental and Evaluative Performance Appraisal Uses. *Journal of Business and Psychology*, 16: 391-412.
- Burnham, Kenneth P. and David R. Anderson. 2002. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. 2<sup>nd</sup> ed. New York: Springer-Verlag.
- Butler, Ruth. 1987. Task-Involving and Ego-Involving Properties of Evaluations: Effects of Different Feedback Conditions on Motivational Perceptions, Interest, and Performance. *Journal of Educational Psychology*, 79(4): 474-482.
- Butler, Ruth and Mordecai Nisan. 1986. Effects of No Feedback, Task-Related Comments, and Grades on Intrinsic Motivation and Performance. *Journal of Educational Psychology*, 78(3): 210-216.
- Callahan, Judith Scully, Amy L. Brownlee, Mary D. Brtek and Henry L. Tosi. 2003. Examining the unique effects of multiple motivational sources on task performance. *Journal of Applied Social Psychology*, 33: 2515-2535.
- Chan, Joanne C. Y. and Shui-fong Lam. 2010. Effects of different evaluative feedback on students' self-efficacy in learning. *Instructional Science*, 38(1): 37-58.
- Cherry, Todd L. and Larry V. Ellis. 2005. Does Rank-Order Grading Improve Students Performance? Evidence from a Classroom Experiment. *International Review of Economics Education*, 4(1): 9-19.
- Cohen-Charash, Yoshi and Paul E. Spector. 2001. The Role of Justice in Organizations: A Meta-Analysis. *Organizational Behavior and Human Decision Processes*, 86(2), 278-321.

- Colquitt, Jason A. 2001. On the Dimensionality of Organizational Justice: A Construct Validation of a Measure. *Journal of Applied Psychology*, 86(3): 386-400.
- Colquitt, Jason A., Donald E. Conlon, Michael J. Wesson, Christopher O. L. H. Porter and K. Yee Ng. 2001. Justice at the Millennium: A Meta-Analytic Review of 25 Years of Organizational Justice Research. *Journal of Applied Psychology*, 86(3): 425-445.
- Covic, Tanya and Mairwen K. Jones. 2008. Is the essay resubmission option a formative or a summative assessment and does it matter as the grades improve? *Assessment & Evaluation in Higher Education*, 33(1): 75-85.
- Deci, Edward L., James P. Connell and Richard M. Ryan. 1989. Self-Determination in a Work Organization. *Journal of Applied Psychology*, 74(4): 580-590.
- Deci, Edwin L. and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.
- Deci, Edward L. and Richard M. Ryan. 2000. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist*, 55(1): 68-78.
- Deci, Edward L., Richard M. Ryan and Richard Koestner. 1999. A Meta-Analytical Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. *Psychological Bulletin*, 125(6): 627-668.
- Deci, Edward L., Robert J. Vallerand, Luc G. Pelletier and Richard M. Ryan. 1991. Motivation and Education: The Self-Determination Perspective. *Educational Psychologist*, 26 (3 & 4): 325-346.
- Duffy, Kirt E. and Robert E. Webber. 1974. On "Relative" Rating Systems. *Personnel Psychology*, 27: 307-311.
- Dysvik, Anders and Bård Kuvaas. 2008. The relationship between perceived training opportunities, work motivation, and employee outcomes. *International Journal of Training and Development*, 12(3): 138-157.
- Early, Christopher P., Gregory B. Northcraft, Cynthia Lee and Terry R. Lituchy. 1990. Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal*, 33(1): 87-105.
- Farndale, Elaine, Veronica Hope-Hailey and Clare Kelliher. 2011. High commitment performance management: the roles of justice and trust. *Personnel Review*, 40(1): 5-23.
- Field, Andy. 2009. *Discovering Statistics Using SPSS*. 3<sup>rd</sup> ed. London: Sage.
- Gagné, Maryléne and Edward L. Deci. 2005. Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26(4): 331-362.
- Garcia, Stephen M., and Avishalom Tor. 2007. Rankings, standards, and competition: Task vs. scale comparisons. *Organizational Behavior and Human Decision Processes*, 102 (1): 95-108.
- Goodman, Jodi S., Robert E. Wood and Margaretha Hendrickx. 2004. Feedback Specificity, Exploration, and Learning. *Journal of Applied Psychology*, 89(2): 248-262.
- Grote, Dick. 2005. *Forced Ranking: Making performance management work*. Boston, MA: Harvard Business School Press.

- Guralnik, Orna, Eyal Rozmarin and Anne So. 2004. Forced Distribution: Is It Right For You? *Human Resource Development Quarterly*, 15(3): 339- 345.
- Hair, Joseph F. Jr., William C. Black, Barry J. Babin and Rolph E. Anderson. 2010. *Multivariate Data Analysis: A Global Perspective*. 7<sup>th</sup> ed. US: Pearson.
- Hattie, John and Helen Timperley. 2007. The Power of Feedback. *Review of Educational Research*, 77(1): 81-112.
- Hazels Beth and Craig M. Sasse. 2008. Forced Ranking: A review. *SAM Advanced Management Journal*, 73(2): 35-39.
- Hox, Joop J. 2010. *Multilevel Analysis: Techniques and Applications*. 2<sup>nd</sup> ed. New York: Routledge.
- Ilgen, Daniel R., Cynthia D. Fisher and M. Susan Taylor. 1979. Consequences of Individual Feedback on Behavior in Organizations. *Journal of Applied Psychology*, 64(4): 349-371.
- Judge, Timothy and Gerald R. Ferris. 1993. Social Context of Performance Evaluation Decisions. *Academy of Management Journal*, 36(1): 80-105.
- Kim, Jay S. and W. Clay Hamner. 1976. Effect of performance feedback and goal setting on productivity and satisfaction in an organizational setting. *Journal of Applied Psychology*, 61(1): 48-57.
- Kluger, Avraham N. and Angelo DeNisi. 1996. The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, 119(2): 254-284.
- Kluger, Avraham N. and Angelo DeNisi. 1998. Feedback Interventions: Toward the Understanding of a Double-Edged Sword. *Current Directions in Psychological Science*, 7(3): 67-72.
- Kuvaas, Bård. 2003. Employee ownership and affective organizational commitment: employees' perceptions of fairness and their preference for company shares over cash. *Scandinavian Journal of Management*, 19: 193-212.
- Kuvaas, Bård. 2006. Work performance, affective commitment, and work performance: the roles of pay administration and pay level. *Journal of Organizational Behavior*, 27: 365-385.
- Kuvaas, Bård. 2007. Different relationships between perceptions of developmental performance appraisal and work performance. *Personnel Review*, 36(3): 378-397.
- Kuvaas, Bård and Anders Dysvik. 2009. Perceived investment in employee development, intrinsic motivation and work performance. *Human Resource Management Journal*, 19(3): 217-236.
- Larson Jr., James R. 1989. The Dynamic Interplay Between Employees' Feedback-Seeking Strategies and Supervisors' Delivery of Performance Feedback. *Academy of Management Review*, 14(3): 408-422.
- Lawler, Edward E. III. 2002. The Folly of Forced Ranking. *Strategy + Business*, 28: 28-32.
- Lind, Allan, Ruth Kanfer and P. Christopher Early. 1990. Voice, Control and Procedural Justice: Instrumental and Noninstrumental Concerns in Fairness Judgments. *Journal of Personality and Social Psychology*, 59(5): 952-959.

- Locke, Edwin A. and Gary P. Latham. 2002. Building a Practically Useful Theory of Goal Setting and Task Motivation. *American Psychologist*, 57(9): 705-717.
- Lucas, Todd. 2009. Justifying Outcomes Versus Processes: Distributive and Procedural Justice Beliefs as Predictors of Positive and Negative Affectivity. *Current Psychology*, 29(4): 249-265.
- Meisler, Andy. 2003. Dead Man's Curve. *Workforce*, 82: 44-49.
- Meyer, Herbert H. 1975. The Pay for Performance Dilemma. *Organizational Dynamics*, 3: 39-50.
- Olson C.A., and Davis G.M. 2003. Pros and cons of forced ranking and other relative performance rating systems. *Society for Human Resource Management Legal Report*, March.
- Pallant, Julie. 2010. *SPSS: Survival Manual*. 4<sup>th</sup> ed. Maidenhead: McGraw-Hill: Open University Press.
- Pedhazur, J Elazar and Liora Pedhazur Schmelkin. 1991. *Measurement, Design, and Analysis: An Integrated Approach*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Pfeffer, Jeffrey and Robert I. Sutton. 2006. Evidence-Based Management. *Harvard Business Review*, 84(1), 2006, 63-74.
- Piccolo, Ronald F. and Jason A. Colquitt. 2006. Transformational Leadership and Job Behaviors: The Mediating Role of Core Job Characteristics. *Academy of Management Journal*, 49(2): 327-340.
- Podsakoff, Philip M. and Jiing- Lih Farh. 1989. Effects of Feedback Sign and Credibility on Goal Setting and Task Performance. *Organizational Behavior and Human Decision Processes*, 44: 45-67.
- Roch, Sylvia, Angela M. Sternburgh and Pat M. Caputo. 2007. Absolute vs Relative Performance Rating Formats: Implications for fairness and organizational justice. *International Journal of Selection and Assessment*, 15(3): 302-316.
- Ryan, Richard M. 1982. Control and Information in the Intrapersonal Sphere: An Extension of Cognitive Evaluation Theory. *Journal of Personality and Social Psychology*, 43(3): 450-461.
- Salas, Eduardo, Jessica L. Wildman, and Ronald F. Piccolo. 2009. Using Simulation-Based Training to Enhance Management Education. *Academy of Management Learning & Education*, 8 (4): 559-573.
- Schaufeli, Wilmar B., Arnold B. Bakker and Willem van Rhenen. 2009. How changes in job demands and resources predict burnout, work engagement, and sickness absenteeism. *Journal of Organizational Behavior*, 30: 893-917.
- Schleicher, Deidra J., Rebecca A. Bull and Stephen G. Green. 2009. Rater Reactions to Forced Distribution Rating Systems. *Journal of Management*, 35(4): 899-927.
- Schunk Dale H., Paul R. Pintrich and Judith L. Meece 2008. *Motivation in Education: Theory, Research and Applications*. 3<sup>rd</sup> ed. New Jersey: Pearson Education, Inc.
- Scullen, Steven E., Paul K. Bergey and Lynda Aiman-Smith. 2005. Forced Distribution Rating Systems and the Improvement of Workforce Potential: A Baseline Simulation. *Personnel Psychology*, (58): 1-32.

- Sears David, and Don McDermott. 2003. The Rise and Fall of Rank and Yank. *Information Strategy: The Executive's Journal*, 19: 6-11.
- Simons, Tony and Quinetta Roberson. 2003. Why Managers Should Care About Fairness: The Effects of Aggregate Justice Perceptions on Organizational Outcomes. *Journal of Applied Psychology*, 88(3): 432-443.
- Singer, Judith D. 1998. Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models. *Journal of Educational and Behavioral Statistics*, 24(4): 323-355.
- Smith, P. and D. Beaton. 2008. Measuring change in psychosocial working conditions: methodological issues to consider when data are collected at baseline and one follow-up time point. *Occupational and Environmental Medicine*, 65(4): 288-296.
- Stevens, James P. 2009. *Applied Multivariate Statistics for the Social Sciences*. 5<sup>th</sup> ed. New York: Routledge.
- Tabachnick, Barbara G. and Linda S. Fidell. 2007. *Using Multivariate Statistics*. 5<sup>th</sup> ed. US: Pearson.
- Taras, Maddalena. 2005. Assessment – Summative and Formative – Some Theoretical Reflections. *British Journal of Educational Studies*, 53(4): 466-478.
- West, Brady T. 2009. Analyzing Longitudinal data With the Linear Mixed Models Procedure in SPSS. *Evaluation & the Health Profession*, 32(3): 207-228.
- West, Brady T., Kathleen B. Welch & Andrzej T. Galecki. 2007. *Linear Mixed Models: A Practical Guide Using Statistical Software*. Boca Raton Fla: Chapman & Hall.

## Appendices

### ***APPENDIX A – Measures***

#### **Intrinsic Motivation**

5 point Likert - 1= strongly disagree, 5 = strongly agree (Kuvaas & Dysvik 2009).

This section contain 6 items describing the inner drive to accomplish your tasks:

1. The tasks that I did in the simulation were themselves representing a driving power
2. The tasks that I did in the simulation was enjoyable
3. I felt that the simulation was meaningful
4. The simulation was very exciting
5. The simulation was so interesting that it was a motivation in itself
6. I was so inspired by the simulation that I almost forgot everything else around me

#### **Distributive Justice**

5 point Likert - 1= strongly disagree, 5 = strongly agree (adapted from Colquitt 2001).

The following items refer to your feedback score. To what extent:

1. Does your feedback score reflect the effort you have put into the simulation activity?
2. Is your feedback score appropriate for the activity you have completed?
3. Does your feedback score reflect what you have contributed with in the simulation activity?
4. Is your feedback score justified, given your performance?

**Procedural Justice**

5 point Likert - 1= strongly disagree, 5 = strongly agree (adapted from Colquitt 2001).

The following items refer to the procedures to arrive at your feedback score. To what extent:

1. Do you feel that you have been able to express your views and feelings during the procedures used to arrive at your feedback score?
2. Do you feel that you have had any influence over the feedback outcome arrived at by those procedures?
3. Do you think those procedures have been applied consistently?
4. Do you think those procedures have been free of bias?
5. Do you think those procedures have been based on accurate information?
6. Do you feel that you have been able to appeal the feedback outcome arrived at by those procedures?
7. Do you think those procedures have upheld ethical and moral standards?



***APPENDIX B – Control Variables***

**Military experience.** Four items measured military experience. Participants were asked to indicate their highest level of military education, and to answer three items assessing their experience, training and participation in crisis management. The Kaiser-Meyer- Olkin value for the factor analysis of these four items was .513, which is below the recommended value of .6 (Pallant 2010). Nonetheless, Bartlett’s Test reached statistical significance ( $p = .000$ ). Due to the statistical significance of the Bartlett’s test and that the KMO value was not far off the recommended value, a factor analysis was performed. The principal component analysis with oblique rotation (Direct Oblimin) revealed the presence of two factors above Kaiser’s criterion of 1, explaining 75 % of the variance in combination. An examination of the scree plot also supported retaining two factors. The pattern matrix and communalities is presented in Table 9.

An inspection of the pattern matrix reveals that the two-factor solution does not conform to a simple structure with all factor loadings loading strongly on only one component (Pallant 2010). In particular, the item assessing training in crisis management poses a problem because it cross – loads and in addition loads below the recommended value of .6 for our sample size (Hair, Black, Babin & Anderson 2010). As such, it is a candidate for deletion. However, the item is retained because; 1) the communality of the item is .545, suggesting that at least half of the variance of the item is represented by the factor solution, 2) this item is not easily dropped because training in crisis management is deemed to be an important part of military experience, 3) the factor loading of .527 is not far from the recommended value of .6.

**Table 9**

*Pattern matrix and Communalities for Principal Component Analysis with Oblimin Rotation of Two Factor Solution of Military Experience<sup>a</sup>*

Item	Pattern coefficients		Communalities
	1	2	
Military education	-.057	<b>.931</b>	.854
Crisis training	<b>.527</b>	.445	.545
Crisis experience	<b>.904</b>	-.338	.841
Crisis participation	<b>.823</b>	.192	.761

Extraction Method: Principal Component Analysis.

a. Rotation converged in 10 iterations.

There was a weak correlation between the two factors ( $r = .148$ ). The results suggest the presence of two factors. The first factor reflects crisis experience and the second military education. Cronbach's alpha for the crisis experience factor was just below the recommended value of .7 (Hair, Black, Babin & Anderson 2010), with a value of .667. As the military education factor is made up of only one item, internal consistency reliability cannot be estimated.

**APPENDIX C – Questionnaire, Control variables****Q1**

Gender

- Male
- Female

**Q2**

Age .....

**Q3**

Nationality

- Norwegian
- Other

**Q4**

First language

- Norwegian
- English
- Other

**Q5**

What is your present occupation?

- 1st year Bachelor student at BI Norwegian School of Management
- 2nd year Bachelor student at BI Norwegian School of Management
- 3rd year Bachelor student at BI Norwegian School of Management
- 1st year Master student at BI Norwegian School of Management
- 2nd year Master student at BI Norwegian School of Management
- Executive student at BI Norwegian School of Management
- Doctoral student at BI Norwegian School of Management
- Other, please specify \_\_\_\_\_

**Q6**

What is your level of military education (highest level completed)?

- None
- Initial training (Førstegangstjeneste)
- Officer training (Befalsskole-/kurs)
- Military academy (Krigsskole)
- Higher (Stabsskole eller høyere)

**Q7**

Do you have any experience with strategic games (computer games, board games, card games etc)?

- None
- Some
- Extensive

**Q8**

To what extent have you been trained in crisis management (civilian or military)?

- None
- Some
- Extensive

**Q9**

To what extent do you have real experience from crisis management (civilian or military)?

- None
- Some
- Extensive

**Q10**

Have you participated in crisis management exercises (civilian or military)?

- None
- One or two exercises
- Three or more exercises

**APPENDIX D – The tasks****Scenario 1**

<b>TaskID</b>	<b>Level</b>	<b>Info</b>
1	Init	Emergency Central reports a car crash on highway E134 close to Odda in Hordaland. Police unit and ambulance have been dispatched to the site of the accident.
1	Early	The police unit at the accident site close to Odda reports that at least two people are severely injured and need immediate transportation to hospital. A rescue helicopter is requested for immediate assistance.
1	Mature	The police unit at the accident site near Odda reports that the conditions of the two injured people is critical, and repeat the request for rescue helicopter.
1	Closed	The rescue operation failed as it was not executed on time.
2	Init	Emergency Central reports a car crash on highway 7 close to Gol in Buskerud. Police unit and ambulance have been dispatched to the site of the accident.
2	Early	The police unit at the accident site close to Gol reports that at least two people are severely injured and need immediate transportation to hospital. A rescue helicopter is requested for immediate assistance.
2	Mature	The police unit at the accident site near Gol reports that the conditions of the two injured people is critical, and repeat the request for rescue helicopter.
2	Closed	The rescue operation failed as it was not executed on time.
3	Init	Emergency central reports a crash on highway E39 close to Stryn in Sogn og Fjordane. One passenger bus and at least two private cars are involved in the crash. Police units and ambulances have been dispatched to the site of the accident. The crash has caused long queues of cars in both directions on the highway.
3	Early	The police units at the accident site near Stryn report that more than 20 people may be injured and some of them severely. All available helicopters are requested for immediate medical assistance and transportation to hospital. Traffic has been redirected and space has been cleared for helicopter landing.
3	Mature	The police units at the accident site near Stryn report that the conditions of many injured people are critical, and repeat the request for all available helicopters for rescue and transportation to hospital.
3	Closed	The rescue operation failed due to low capacity.
4	Init	Emergency central has received SOS signals from a sailing boat that has capsized 5 NM south of Risør in Aust-Agder. A boat from the local police has been dispatched to investigate.
4	Early	The police boat has arrived at the location of the capsized sailing boat outside Risør, and report that there are more than 10

	people floating in the sea. Some of them do not have swimwear, and appear to be exhausted. There is no sign of the sailing boat. Rescue helicopter is requested for immediate assistance.
4 Mature	The police boat outside Risør reports that the situation is critical for the people that are floating in the sea, and repeats the request for immediate assistance from rescue helicopter.
4 Closed	The rescue operation failed due to low capacity.
5 Init	Emergency central reports that the flood that has been building up gradually in the Glomma river close to Kongsvinger (Hedmark) has caused several houses to be trapped in flood water between river banks.
5 Early	The flood at Kongsvinger is increasing, and more than 10 houses are flooded. There is an urgent need to provide transportation of people that have been trapped inside their houses.
5 Mature	People are standing on rooftops and getting more and more desperate to be transported out of the flood area. The emergency central calls for all available transportation resources to relieve the people in distress.
5 Closed	The rescue operation failed due to low capacity.
6 Init	The Military Command Center reports that a foreign fishing vessel suspected of illegal fishing has been located to the west of Florø (Sogn og Fjordane). The vessel needs to be identified and kept under surveillance from the air as long as it is within proximity of Norwegian waters.
6 Early	The fishing vessel has been identified as MS Ramona of Portugal. It has previously been escorted out of Norwegian waters due to illegal fishing. In addition to surveillance, the Coast Guard requests helicopter support for transporting inspectors on board the vessel.
6 Mature	The Coast Guard have now made visual and radio contact with MS Ramona, and the suspicion of illegal fishing has been strengthened. The Coast Guard repeats its request for surveillance and helicopter support, and stress the urgency.
6 Closed	You did not succeed in the security mission
7 Init	A Norwegian fishing vessel has reported to the Military Command Center that a submarine of unknown nationality has been detected as it surfaced and later submerged to the west of Smøla (Møre og Romsdal). The submarine needs to be identified and kept under surveillance as long as it is within proximity of Norwegian waters.
7 Early	Based on visual indicators it is highly probable that the submarine is Russian and belonging to the Kursk class. The Military Command Center has just received images of a Tupolev long-range bomber aircraft which is circling within visual distance of the submarine.
7 Mature	The Military Command Center increases its alert level as it appears that the Russian aircraft is changing course and heading

		directly for Norwegian mainland. Your orders are to maintain surveillance while immediately effectuating air intercept of the Russian aircraft.
7	Closed	You did not succeed in the security mission.
8	Init	The Military Command Center has received several SOS signals from the Russian submarine located outside Smola.
8	Early	The SOS signals from the Kursk class submarine are becoming more frequent. Visual contact has been made, and it seems that the submarine crew have launched several lifeboats.
8	Mature	There are more than three lifeboats in the sea close to where the Russian submarine was last detected. In addition, more than 30 people are reported floating in the sea. Storm in the area will make a rescue operation difficult.
8	Closed	You did not succeed in the security mission.
9	Init	The Military Command Center has received intelligence that a terrorist attack is being planned towards a vital military installation close to Rena military camp, north of Elverum (Hedmark). The attack is being executed within the next 45 minutes.
9	Early	Norwegian special forces have located a vehicle close to Elverum, with confirmed terrorists on board. The vehicle is heading towards the Rena military camp.
9	Mature	The terrorist vehicle approaches the Rena military camp loaded with explosives. An attack seems imminent.
9	Closed	You did not succeed in the security mission.
10	Init	The Military Command Center has received intelligence that a terrorist attack is being planned towards the nuclear reactor in Halden (Ostfold). The attack is being executed within the next 45 minutes.
10	Early	Norwegian special forces have located a vehicle close to Halden, with confirmed terrorists on board. The vehicle is heading towards the Halden nuclear reactor.
10	Mature	The terrorist vehicle approaches the Halden nuclear reactor loaded with explosives. An attack seems imminent.
10	Closed	You did not succeed in the security mission.
11	Init	An oil worker on board the platform Safe Security, located to the south of Shetland, has been brought to Stavanger for police questioning after she allegedly made statements indicating that there might be a bomb hidden on board the platform. Her motives for making such a statement are not known.
11	Early	Norwegian police, supported by military special forces, have started searching the platform for explosive devices. The police requests helicopters to stand by for possible evacuation.
11	Mature	The search for explosive devices on board the platform Safe Security continues. Nothing is found so far. There does not seem to be need for evacuation.
11	Closed	The entire platform was searched, and nothing suspicious found

- false alarm!

12	Init	The Hordaland Police Department has received SOS messages from the cruise ship MS Bergen that there have been explosions on board, possibly triggered by sabotage from terrorists. The ship has lost its steering and is taking in water. The ship is drifting towards Norwegian mainland and the Bergen area. The crew have started evacuation of the 500 passengers on board.
12	Early	The evacuation of the ship MS Bergen continues. It is clear that there are too few life boats to accommodate all the passengers and crew. In addition, fire breaks out both on board and in the oil spillage close to shore.
12	Mature	The shipwreck situation outside Bergen becomes increasingly critical as there are reports of people drowning due to lack of lifeboat capacity. Nearby ships are reluctant to enter the area due to burning oil spillage.
12	Closed	The rescue and fire fighting did not succeed due to low capacity.

## Scenario 2

TaskID	Level	Info
1	Init	Emergency Central reports a crash between a motorbike and a trailer on E16 close to Voss in Hordaland. Police unit and ambulance have been dispatched to the site of the accident.
1	Early	The police unit at the accident site close to Voss reports that the biker is unconscious and need to get to the hospital as soon as possible.
1	Mature	The police unit at the accident site near Voss reports that the conditions of the biker is critical, and repeats the request for a rescue helicopter.
1	Closed	The rescue operation failed as it was not executed on time.
2	Init	Emergency Central reports that a car has been observed ten meters of the road near E6 close to Dovre (Oppland). At least one injured person is observed inside the car wreck. Police unit and ambulance have been dispatched to the site of the accident.
2	Early	The police unit at the accident site close to Dovre reports that at least two people are severely injured and need immediate transportation and rescue assistance.
2	Mature	The police unit at the accident site near Dovre reports that the conditions of the two injured people are critical, and repeats the request for rescue assistance.
2	Closed	The rescue operation failed as it was not executed on time.
3	Init	Emergency central reports a crash on highway 13 close to Odda in Hordaland. One truck and 4 private cars are involved in the crash. Police units and ambulances are on their way.
3	Early	The police units at the accident site near Odda report that at least 18 people are involved in the crash, and some people's



---

	conditions are critical. All available helicopters are requested for immediate medical assistance and to bring injured people to hospital. Traffic has been redirected and space has been cleared for helicopter landing.
3 Mature	The police units at the accident site near Odda stress that the conditions of several injured people are critical, and repeat the request for all available helicopters for rescue and transportation to hospital.
3 Closed	The rescue operation failed due to low capacity.
4 Init	Emergency central has received SOS signals from a rowboat that has overturned 5 NM west of Sandnes in Rogaland. A boat from the local police is on its way to investigate the situation.
4 Early	The police boat has arrived at the location of the overturned rowboat outside Sandnes, and report that there are 6 people floating in the sea and some of them are unconscious.
4 Mature	Rescue helicopter is requested for immediate assistance.
4 Closed	The rescue operation failed due to low capacity.
5 Init	Emergency central reports that Orkla river close to Orkdal (Sor Trondelag) is flooding over. More than 15 houses might be trapped in the water between river banks.
5 Early	The flood at Orkdal is increasing. Some houses are in danger of being severely damaged. Many people are not able to get out of their houses, and some are standing on rooftops screaming for help. There is an urgent need to transport people out of the area. Emergency central has received information about two hikers and one fishing man in the area, who are not localized.
5 Mature	There is a concern that they might be caught by the water. The emergency central calls for all available transportation and rescue resources to help people in need.
5 Closed	The rescue operation failed due to low capacity.
6 Init	The Military Command Center reports that a vessel with an unknown mission has been located to the west of Stavanger (Rogaland). Other fishing vessels in the area are concerned about its intentions. The vessel needs to be identified and kept under surveillance as long as there is insecurity regarding its mission in Norwegian waters.
6 Early	The observed vessel is not yet identified, and there is suspicion that it is cooperating with a second vessel that is observed to hold the same course. The request for surveillance is expanded, as both vessels need to be identified.
6 Mature	The Coast Guard have now made visual and radio contact with one of the boats, which is identified as MS Ramon. It has previously been escorted out of Norwegian waters due to illegal fishing. In addition to surveillance, the Coast Guard now requests helicopter support for transporting inspectors on board this vessel. The other vessel is still not identified, and the request for surveillance stands.

---

6	Closed	You did not succeed in the security mission
7	Init	The Military Command Center receives information from several fishing vessels indicating that there are 2 submarines of unknown nationality in the area. These have been detected as they surfaced and later submerged in the area west of Heroy (Nordland). The 2 submarines need to be identified and kept under surveillance as long as they are within proximity of Norwegian waters.
7	Early	Based on visual indicators it is highly probable that the 2 submarines are Russian and belonging to the Kursk class. The Military Command Center has just received images of a Tupolev long-range bomber aircraft which is circling within visual distance of the submarines.
7	Mature	The Military Command Center increases its alert level as it appears that the Russian aircraft is changing course and heading directly for Norwegian mainland. Your orders are to maintain surveillance of the submarines while immediately effectuating air intercept of the Russian aircraft.
7	Closed	You did not succeed in the security mission.
8	Init	The Military Command Center has received frequent SOS signals from the Russian submarine located west of Bomlo (Hordaland).
8	Early	Visual contact has been made, confirming that the submarine crew has launched several lifeboats.
8	Mature	There are more than four lifeboats in the sea beside the Russian submarine. In addition, more than 50 people are reported floating in the sea. They all seem to be exhausted and a few are in shock having problem swimming. The storm has decreased in intensity, so a rescue operation should not be too difficult.
8	Closed	You did not succeed in the security mission.
9	Init	The Military Command Center has received intelligence that a bomb has exploded resulting in a fire at a vital military installation at Oerland air force base, north of Trondheim (Sor Trondelag).
9	Early	Norwegian special forces have not been able to locate the terrorists.
9	Mature	More than 10 people may be injured. There is an urgent request for all available aircraft that can assist in the fire and bring injured people to hospital.
9	Closed	You did not succeed in the security mission.
10	Init	The Military Command Center has received a message that a sabotage operation will take place at the nuclear reactor at Kjeller (Ostfold), close to Oslo. The attack is being executed within the next 50 minutes.
10	Early	Norwegian special forces have observed two vans close to Oslo, both with confirmed terrorists on board. The two vans are approaching the nuclear reactor from different directions. There is a request for immediate interception of the two vans.

---

		Five police cars are tracing the vans. They repeat the request for
10	Mature	interception since an attack seems imminent.
10	Closed	You did not succeed in the security mission.
11	Init	An oil worker at the platform Gullfaks, located west of Bergen (Hordaland), has been brought to Bergen for police questioning after indicating that there might be explosive devices hidden on board the platform.
11	Early	Norwegian police, supported by military special forces, are searching the platform for explosive devices. The police requests helicopters to stand by in case of evacuation.
11	Mature	There has now been found explosive devices on board the platform. All available resources are needed for evacuation.
11	Closed	The entire platform was searched, and nothing suspicious found - false alarm!
12	Init	The More and Romsdal Police Department has received SOS messages from Hurtigruta outside Kristiansund that there is a fire on board. No one knows what caused the fire. People are panicking and at least 10 persons have jumped over board. The ship is drifting towards Kristiansund. The crew needs to evacuate all of the 450 passengers on board.
12	Early	It is clear that there are 20 lifeboats available for the passengers and the crew, which is not enough to rescue all the people on board. Due to high waves it is difficult for other boats to come to assistance.
12	Mature	The Hurtigruta situation outside Kristiansund becomes more severe as people in the lifeboats and in the water are struggling in the waves. Nearby ships are still reluctant to enter the area due to high waves.
12	Closed	The rescue and fire fighting did not succeed due to low capacity.

Marthe Aune  
Julie Røed

GRA 19002 Preliminary Thesis Report:

# **- Forced Ranking: Friend or Foe? -**

Advisors: Thorvald Hærem and Anders Dysvik

Hand-in date:  
17.01.2011

Place:  
BI Oslo

Program:  
Master of Science in Leadership and Organizational Psychology

---

## Table of Content

<b>Abstract .....</b>	<b>ii</b>
<b>1. Introduction .....</b>	<b>1</b>
<b>1.1. Research Model .....</b>	<b>2</b>
<b>1.2. Outline .....</b>	<b>2</b>
<b>2. Theory and Hypotheses .....</b>	<b>2</b>
<b>2.1 Forced Ranking .....</b>	<b>2</b>
2.1.1. Forced Ranking and Feedback .....	4
2.1.2. Relationship Patterns of Forced Ranking and Performance .....	5
2.1.3. Feedback and Performance .....	6
<b>2.2. Intrinsic Motivation.....</b>	<b>8</b>
<b>2.3. Fairness .....</b>	<b>10</b>
<b>3. Method.....</b>	<b>12</b>
<b>3.1. Experimental Task.....</b>	<b>12</b>
<b>3.2. Participants and Procedure.....</b>	<b>13</b>
<b>3.3. Measures .....</b>	<b>13</b>
3.3.1. Forced Ranking.....	13
3.3.2. Intrinsic Motivation (M) .....	13
3.3.3. Fairness (M).....	13
3.3.4. Performance (Y) .....	13
<b>Appendices .....</b>	<b>14</b>
<b>APPENDIX A.....</b>	<b>14</b>
<b>Reference list.....</b>	<b>15</b>
<b>Articles.....</b>	<b>15</b>
<b>Books .....</b>	<b>18</b>

***Abstract***

For our master thesis we want to investigate forced ranking. We have looked into theory and previous research on forced ranking, feedback, intrinsic motivation and fairness and on this note formulated hypotheses. In the method section we present the procedure of the experiment, and suggest appropriate measures.

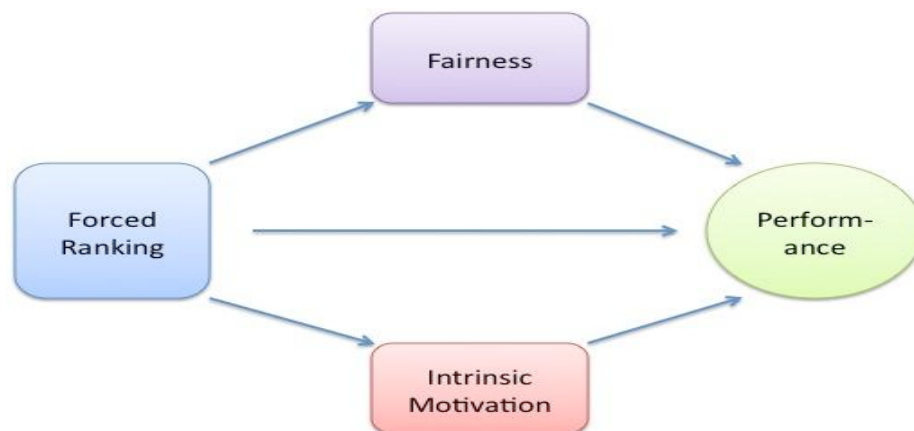
## 1. Introduction

Most, if not all, organizations depend on performance in order to survive (Grant 2009). Knowledge on what influence performance is thus immensely important. An important factor that has been identified is feedback. Feedback is an inevitable part of all organizations as it is a form of communication that conveys some degree of information about past behavior, performance or achieved understanding (Ilgen, Fisher and Taylor 1979). Hattie and Timperley (2007) recognize feedback as one of the most powerful influences on achievement comparing it with factors such as prior cognitive ability, and socioeconomic influences. Naturally there have also been devoted a lot of research to this field. However, there is one area within this field that has stayed away from the academic limelight, namely feedback in the form of forced ranking (Scullen, Bergey and Aiman-Smith 2005; Blume, Baldwin and Rubin 2009; Schleicher, Bull and Green 2009).

Forced ranking represents a type of performance appraisal and refer to the procedure of distinguishing individuals into preexisting performance categories (Olson & Davis 2003). As such, forced ranking by definition categorizes people on a relative scale. Forced ranking has stirred quite a debate, with strong opinions about the possible detrimental and beneficial effects of forced ranking on both sides. Despite anecdotal evidence, there is a lack of empirical research investigating the effects of forced ranking on performance. The purpose of this paper is thus to shed some light on this relationship, and by doing so also add to the general feedback literature. Moreover, in order to be able to better explain the relationship between forced ranking and performance, we identify intrinsic motivation and fairness as two potentially mediating variables. On this basis we propose the following research question:

*How does forced ranking affect intrinsic motivation, fairness and individual performance?*

### **1.1. Research Model**



### **1.2. Outline**

The paper adheres to the following outline. First, we will present anecdotal knowledge and relevant research on forced ranking. Then we will provide theory and research on feedback, intrinsic motivation and fairness, and finally present our methodology and measures of interest.

## **2. Theory and Hypotheses**

In this section we review research and theories on forced ranking, feedback, intrinsic motivation and fairness. On this basis we suggest hypotheses.

### **2.1 Forced Ranking**

Performance evaluation systems are one of the most frequently used human resource management systems in organizations today (Blume, Baldwin & Rubin 2009). However, despite their pervasive use, previous research has identified several problems in relation to the application of performance evaluations, notably rating errors stemming from rater bias. The bias devoted considerably attention is the tendency on part of the raters to give lenient or inflated ratings. Consequently, this bias results in a lack of differentiation between high and low performers (Blume, Baldwin & Rubin 2009). Typically, ratings become constricted so that more than 90% of the evaluations are distributed between the highest



and next to highest-ratings (Guralnik, Rozmarin & So 2004). Given such inflation, it is argued that performance evaluation systems lose their credibility, as they do not differentiate between the employees (ibid). Recently, it has therefore been a revival of forced ranking systems, which were developed to deal with the leniency bias and lack of differentiation. Despite limited evidence linking forced ranking to actual improved organizational performance, the use of forced ranking in organizations proliferated greatly (Blume, Baldwin & Rubin 2009). In fact, recent estimates are that approximately one-fifth of *Fortune* 1000 companies use some form of forced ranking systems (Sears & McDermott 2003).

Forced ranking is a type of performance appraisal where evaluations are required to fit along the lines of a particular distribution (Schleicher, Bull and Green 2009). This performance evaluation approach is based on the repeated finding in social sciences that when measured in large enough samples, most human phenomena tend to follow a normally distributed curve (Guralnik, Rozmarin & So 2004). The “archetype” of forced ranking is thus the procedure of categorizing individuals into preexisting performance categories, against other employees in the department or peer group (e.g. a 20-70-10 distribution) (Olson & Davis 2003; Hazels & Sasse 2008). These performance rankings are then applied to a bell curve, with those ranking at the bottom (usually 10%) being put on probation, given improvement possibilities or terminated. By contrast, those ranking on top (usually 20%) are generously rewarded for their performance (Hazels & Sasse 2008). Obviously, forced ranking systems contrast with an absolute system of evaluation, in which employees are evaluated with basis in an absolute standard and not in relation to other ratees (Duffy & Webber 1974).

The recent revival of forced ranking has not gone unnoticed. Forced ranking has been and still is a controversial issue, resulting in a heated debate over the pros and cons in both professional HR journals and the media (e.g. Meisler 2003). Jack Welch, the former superstar CEO of General Electric, is one of the most famous proponents of forced ranking (or the “vitality curve” as referred to in the Welchian lingo), arguing that forced ranking is the key to the organization’s

competitive advantage (Schleicher, Bull & Green 2009). Nonetheless, respected authors have questioned the validity and effectiveness of the forced ranking approach (Pfeffer & Sutton 2006). It is perhaps not very difficult to imagine why forced ranking has stirred such a debate. While traditional performance appraisals usually are criterion-based (establishing a performance level), forced ranking is about distinguishing people (Hazels & Sasse 2008). Thus, ratings may be perceived to be unfair because managers are forced to place their employees into categories based on a normally distributed curve, regardless of actual performance is normally distributed (Roch, Sternburgh & Caputo 2007).

### ***2.1.1. Forced Ranking and Feedback***

Clearly, performance ratings such as forced ranking provide an important source of feedback to individuals in organizations (Bartol, Durham & Poon 2001). Despite impassioned anecdotal accounts (e.g. Lawler 2002; Grote 2005) on both side of the debate, very little empirical research has emerged on forced ranking (Blume, Baldwin & Rubin 2009). Therefore, it is of interest to look into what the feedback literature could contribute with to this debate. On a general level, feedback is a form of communication that conveys some degree of information about past behavior, performance or achieved understanding (Ilgen, Fisher and Taylor 1979; Hattie and Timperley 2007). Yet having this stated it is important to note that feedback is far from a simple stimulus as feedback has several dimensions. One of the most important distinctions is feedback sign- whether the feedback is positive or negative (Ilgen, Fisher & Taylor 1979, Podsakoff and Farh 1989). Furthermore, feedback can be conceptualized as formative or summative and norm-referenced or self-referenced (Chan & Lam 2010). Summative feedback focuses on the outcome, whereas formative feedback provides the individual with learning cues in how to progress (Taras 2005; Covic & Jones 2008). Self-referenced feedback involves self-comparative appraisal, whereas norm-referenced concerns social-comparative appraisal (Chan & Lam 2010). Applied to the case of forced ranking it seems reasonable to argue that forced ranking is a summative, norm-referenced form of feedback. The feedback sign will depend on in what category the individual's performance is deemed

appropriate to be placed in relative to others in the peer group. Thus, forced ranking is summative feedback that is distributed according to a variant of the normal distribution (e.g. 20-70-10).

### ***2.1.2. Relationship Patterns of Forced Ranking and Performance***

The general little attention that has been devoted to the field of forced ranking also transmits to our knowledge on the relation between forced ranking and performance. Nonetheless, there are some studies that have approached the issue of forced ranking. A simulation study conducted by Scullen, Bergey and Aiman-Smith (2005) investigated if implementation of a forced distribution rating system (FDRS) could improve the average quality of an organization's workforce. Their findings revealed that FDRS could in fact improve the workforce potential, however, potential side effects such as decline in employee moral, general dissatisfaction, lowered organizational commitment and possible increase in turnover were identified. Research conducted by Garcia and Tor (2007) provides further knowledge on why these negative effects might occur. By nature, forced ranking involves social comparison. Leaning on research and findings from Festinger, Garcia and Tor (2007) claim that this comparison process often results in competitive behavior. Their findings indicate that it is competition on a general scale rather than task comparison that is the main social comparison facilitator of competitive behavior, and therefore suggest that forced ranking can lead to greater competition among the employees. Greater competition may sound positive; however this can actually have a detrimental effect. As stated by Garcia and Tor (2007, 106):

while highly ranked employees may be more competitive and productive through simple self selection, the championing of forced rankings fails to anticipate how competitive forces may ultimately inhibit the profit-maximizing exchange or pooling of information and resources among those 'star' employees.

On the basis of the study by Scullen et al. (2005) it is possible to argue that there is a relationship between forced ranking and organizational performance. However, we are interested in how forced ranking relates to individual

performance, and therefore this study holds marginally value. As forced ranking is a type of feedback we turn back to the feedback literature.

### ***2.1.3. Feedback and Performance***

It is a well-established finding that feedback is related to performance (Kim & Hamner 1976; Illgen, Fisher & Taylor 1979; Larson Jr. 1989; Early, Northcraft, Lee & Lituchy 1990; Kluger & DeNisi 1998; Goodman, Wood & Hendrickx 2004; Hattie & Timperley 2007; Anseel, Lievens & Schollaert 2009). As pointed to above, forced ranking is a form of performance feedback, hence forced ranking are expected to relate to performance in some way or another. The initial theoretical arguments for the effectiveness of feedback were provided by Thorndike and his law of effect (Kluger & DeNisi 1998). Positive feedback was equated with reinforcement, and negative feedback with punishment. Both types of feedback should improve performance because positive feedback reinforces performance, whereas negative feedback punishes the erroneous behavior (Kluger & DeNisi 1998). An influential and much cited review by Ammons (1956) gave further support for the beneficial effect of feedback on performance. Given this understanding of feedback we would expect that feedback provided through a forced ranking system would be beneficial to performance because the top 20% should be even more motivated to perform, whereas those in the middle and low categories would get a kick in the pants to enhance their performance.

Nevertheless, more recent research on feedback (e.g. Kluger and DeNisi 1996) suggests that not all feedback necessarily lead to better performance. Indeed, the presence of negative effects of feedback is robust; about 34-38 % of the effect sizes investigated in a meta-analysis by Kluger and DeNisi (1996) showed a negative effect on subsequent performance. The theoretical explanation provided (Feedback Intervention Theory) suggests that feedback that directs attention to the self (for example "You are a great student") is more likely to attenuate the effect of feedback on performance. By contrast, feedback effects on performance are augmented by feedback that is related to the task (for

example “This essay can be improved if elaborating more on the theoretical concepts”). The explaining mechanism is that cues that shifts attention to the self reallocates cognitive resources from task to the self, and in such a way weaken performance (Kluger & DeNisi 1996). The major discriminator is thus whether feedback is directed to the task or to the self level (Hattie & Timperley 2007). Moreover, grading research also supports that feedback in form of grades could have a negative effect on performance. Although Cherry and Ellis (2005) found that rank-order grading could generate improved student performance relative to criterion- referenced grading, Butler and Nisan (1986) found that grades might encourage an emphasis on quantitative aspects of learning, reduce creativity, promote fear of failure, and weaken interest. Thus, the picture gets complicated, as the effect of feedback is not as straightforward as was hypothesized in earlier research.

The distinction whether feedback is directed to self or task is useful in relation to forced ranking. Given that feedback derived from forced ranking is normative, that is, feedback that conveys comparative information, it could be argued that this type of feedback diverts attention from the task to the self. Feedback that directs attention to the self via normative cues has been shown to be largely ineffective (Kluger and DeNisi 1998). Similarly, a study by Butler (1987) found that grades increased ego involvement, but did not affect performance relative to the no-feedback control group. This contradicts the understanding of feedback given by Thorndike and Ammons (Kluger & DeNisi 1998), in that feedback is not universally positively linked to performance. Nonetheless, it could also very well be argued that forced ranking feedback also conveys information regarding task performance, although on a relative scale. In contrast to feedback directed to self, task- focused feedback have shown to increase task involvement and consequently performance (Butler 1987). In particular, feedback that provides corrective information (e.g. formative feedback) has been shown to be effective in relation to performance (Hattie & Timperley 2007). Thus, we run into muddy waters, as the forced ranking feedback does not neatly fall into the distinction between feedback directed to task or self. Consequently,

it is somewhat difficult to predict the effects of forced ranking on performance. However, given that forced ranking conveys comparative information, it would be hard to overlook the social comparison aspect of forced ranking, which is argued to attenuate performance (Kluger & DeNisi 1996). Furthermore, as negative feedback is found to be more potent than positive feedback at the self level (Hattie & Timperley 2007), we argue that the most pronounced effects of forced ranking will be seen in the low (10%) category. Coupled with the fact that the forced ranking feedback does not indicate any learning cues (corrective information is limited) it seems likely that the feedback obtained will not have a strong effect on performance in any of the categories. Based on the above account, we therefore hypothesize the following:

***H1a.** Subjects ranked high will maintain or decrease performance after receiving forced ranking feedback.*

***H1b.** Subjects ranked middle maintain or decrease performance after receiving forced ranking feedback.*

***H1c.** Subjects ranked low will decrease performance after receiving forced ranking feedback.*

## **2.2. Intrinsic Motivation**

Owing to the limited research on forced ranking, we know little about what factors that might intervene between forced ranking and performance. However, as forced ranking represents a type of feedback, we expect that the effect of forced ranking on performance should be mediated by motivation. This as feedback is widely believed to affect motivation (Bandura 1993; Deci & Ryan 2000; Locke & Latham 2002), which subsequently affects performance by directing attention and increasing persistence and effort (Callahan et al. 2003). We suggest that intrinsic motivation has the potential to fill this mediating role.

Intrinsic motivation is a core tenant in Deci and Ryan's self-determination theory (SDT) and cognitive evaluation theory (CET). CET was presented by Deci and Ryan

as a sub-theory within SDT with the aim of specifying factors that explain variability in intrinsic motivation (Deci & Ryan 2000). Intrinsic motivation can be defined as the motivation to perform an activity for itself, in order to experience the pleasure and satisfaction inherent in the activity (Deci and Ryan 1985). According to the STD and CET feedback that are interpreted as information about one's competence and satisfy individuals' need for autonomy will enhance intrinsic motivation (Deci, Ryan & Koestner 1999; Gagne & Deci 2005). Moreover, research has shown that positive performance feedback can enhance intrinsic motivation, and that negative performance feedback can diminish it (Deci, Ryan & Koestner 1999; Deci & Ryan 2000;). However, positive feedback that is perceived as controlling, that is, positive feedback having an evaluative character, or emphasizing how one should perform – has clearly shown to decrease intrinsic motivation (Ryan 1982; Deci, Connell & Ryan 1989). Ryan (1982) suggests that positive feedback can be perceived either as informal or controlling depending on various factors, and that these will determine whether the positive feedback increases or decreases intrinsic motivation. Drawing this link to forced ranking one could believe that positive rank could be interpreted as controlling as forced ranking have a normative character and thus state something about how a person should perform leading to a decrease in intrinsic motivation. However, a positive rank could also be interpreted as information about one's competence and therefore increase intrinsic motivation. For participants' receiving a negative rating it is possible to infer that they will experience a decline in intrinsic motivation in both of the conditions – either they perceive the ranking as controlling or as decreasing perceptions of competence: both of which decreasing intrinsic motivation. Accordingly, there is reason to believe that forced ranking, which provides rating dependent on performance relative to others, may operate via intrinsic motivation to influence performance.

Before the millennium few studies had examined the performance effects associated with intrinsic motivation. Rather, much of the intrinsic motivation literature had focused on how extrinsic motivational sources affect intrinsic

motivation (Callahan et al. 2003). Today however we have achieved a great deal more knowledge on this relation. For example, Callahan et al. (2003) examined the unique effects of multiple sources on task performance, and found that intrinsic motivation had the greatest effect on performance. Moreover, two studies by Kuvaas (2006; 2007) reported a positive relation between intrinsic motivation and performance. In a study on transformational leadership and job behaviors Piccolo and Colquitt (2006) also found that the indirect effect of intrinsic motivation supported the direct effect of transformational leadership on task performance. Finally, a study by Dysvik and Kuvaas (2008) observed that the relationship between perceived training opportunities and work performance was fully mediated by intrinsic motivation. All these studies thus suggest that intrinsic motivation is a potent predictor of task performance. Accordingly we hypothesize the following:

***H2. The relationship between forced ranking and performance will be mediated by intrinsic motivation.***

### **2.3. Fairness**

Fairness has been identified as a potentially important variable in the debate over forced ranking (Lawler III 2002; Meisler 2003; Olson & Davis 2003; Sears & McDermott 2003), and a peak into the organizational justice research therefore seems appropriate. Research on organizational justice has identified different forms of justice, most notably distributive justice and procedural justice (Colquitt 2001). Distributive justice refers to the perceived fairness of outcomes, whereas procedural justice refers to the perceived fairness of the process by which outcomes were arrived at (Cohen-Charash & Spector 2001). A number of studies have linked fairness perceptions to important organizational outcomes such as affective commitment (Kuvaas 2003), organizational commitment (Farndale, Hope-Hailey & Kelliher 2011), performance (Lind, Kanfer & Earley 1990) and turnover (Simons & Roberson 2003). Thus, it would be in the best



interest of organizations to maximize employees' fairness perceptions (Roch, Sternburgh & Caputo 2007).

Although considerable research has documented the importance of fairness perceptions in connection to performance evaluation processes in organizations (Bartol, Durham & Poon 2001), there is however no studies to our knowledge that investigates fairness reactions and subsequent performance after receiving forced ranking feedback. Nonetheless, research on performance evaluations has identified the format of performance appraisal systems to be important in connection to fairness. A study by Roch, Sternburgh and Caputo (2007) suggest that relative formats are perceived to be less fair than absolute formats, with the forced ranking format perceived to be the least fair. Furthermore, Bartol, Durham and Poon (2001) link fairness with the segmentation of performance appraisal systems. They point to that a typical three-category system is designed to capture 70-80 % of employees in the middle category. Thus, if performance is normally distributed, an employee receiving performance feedback one standard deviation below average would typically receive the same performance rating as a colleague performing one standard deviation above average, which could influence fairness perceptions.

It is reasonable to argue that both procedural and distributive justice are important in connection to forced ranking. First, equity theory holds that when an individual perceives distributive injustice at work, the employee can alter his or her quality and quantity of work to re-establish justice (Cohen-Charash & Spector 2001). Furthermore, research suggests that people tend to be highly influenced by social comparison information, and that information about an individual's standing within a group influences distributive justice perceptions (Bartol, Durham & Pool 2001). Forced ranking does indeed convey information regarding relative standing within a group; hence it is argued that forced ranking affects the perceptions of distributive justice. For example, given the tendency for people to rate themselves above average (Meyer 1975), it could be argued that people in the low and middle category will perceive distributive justice to be

low. Second, research has shown that when outcomes are low, fairness perceptions of procedural justice becomes more important (Roch, Sternburgh & Caputo 2007). Thus, individuals who receive relatively high ratings may not be particularly concerned about procedural justice, and would be more likely to perceive the ratings as fair. Nonetheless, people are less likely to pay attention to procedural information when people have access to information regarding the outcomes of others (Bartol, Durham & Pool 2001). In the case of forced ranking, although individuals may not have direct information about the performance of others, the categorization still conveys comparative information. For example, if ranked a C player (10<sup>th</sup> percentile) it is obvious that quite a few people have performed better than you. Therefore, because of the availability of comparative information it is argued that the perceptions of distributive and procedural justice will parallel each other. We therefore hypothesize the following:

**H3.** *Fairness perceptions will mediate the relationship between feedback and performance*

**H3a.** *Subjects ranked in the high category will perceive the highest distributive and procedural justice.*

**H3b.** *Subjects ranked in the middle category will perceive less distributive and procedural justice than subjects ranked high.*

**H3c.** *Subjects ranked in the low category will perceive the least distributive and procedural justice.*

### **3. Method**

#### **3.1. Experimental Task**

To investigate our research question and our hypotheses we will use computer simulation. Using this simulation provides many advantages. Maybe the most prominent that it gives a complex model of reality (Salas, Wildman, and Piccolo 2009). This simulation will involve a scenario database, where participants are to be involved in military operations outside the coast of Norway. All of the participants will be provided with a computer with a graphical user interface with

an interactive map of the Norwegian coast. They will receive a number of scenarios and are expected to solve these individually.

### ***3.2. Participants and Procedure***

In order to attain a representative selection we will need 120 participants. We are going to get these from the Master of Science programs, mainly from the master in leadership and organizational psychology. The simulations will start in mid February.

### ***3.3. Measures***

#### ***3.3.1. Forced Ranking***

The participants' will be distinguished into three categories. The categorization will be based on the participants' performance relative to each other. The ranking will moreover have a summative character and we will therefore use a scale from 1 to 3, where 1 represents the best category and 3 the worst. Finally, we will most likely base the categorization on a 20-60-20 distribution.

#### ***3.3.2. Intrinsic Motivation (M)***

Intrinsic motivation is going to be assessed by using a six items measure. The items are descriptive adjectives commonly used to assess intrinsic motivation.

#### ***3.3.3. Fairness (M)***

What fairness measure to use is still under discussion.

#### ***3.3.4. Performance (Y)***

Can be measured on the basis of detection, information search and attack, where each type has a minimum and maximum score. The range of the different performance measures will be divided into three as a basis for the forced ranking.

## Appendices

### ***APPENDIX A***

#### **Intrinsic Motivation**

5 point Likert - strongly disagree/strongly agree (Kuvaas & Dysvik 2009).

This section contain 6 items describing the inner drive to accomplish your tasks:

1. The tasks that I did in the simulation were themselves representing a driving power
2. The tasks that I did in the simulation was enjoyable
3. I felt that the simulation was meaningful
4. The simulation was very exciting
5. The simulation was so interesting that it was a motivation in itself
6. I was so inspired by the simulation that I almost forgot everything else around me

## Reference list

### **Articles**

- Ammons, Robert B. 1956. Effects of knowledge of performance: A survey and tentative theoretical formulation. *Journal of General Psychology*, 54: 279-299.
- Ansel, Frederik, Filip Lievens and Eveline Schollaert. 2009. Reflection as a strategy to enhance task performance after feedback. *Organizational Behavior and Human Decision Processes*, 110: 23-35
- Bandura, Albert. 1993. Perceived Self- Efficacy in Cognitive Development and Functioning. *Educational Psychologist*, 28(2): 117- 148.
- Bartol, Kathryn, Cathy C. Durham and June M. L. Poon. 2001. Influence of Performance Evaluation Rating Segmentation on Motivation and Fairness Perceptions. *Journal of Applied Psychology*, 86(6): 1106-1119.
- Blume, Brian D., Timothy T. Baldwin and Robert S. Rubin. 2009. Reactions to Different Types of Forced Distribution Performance Evaluation Systems. *Journal of Business and Psychology*, (24): 77-91.
- Butler, Ruth. 1987. Task-Involving and Ego-Involving Properties of Evaluations: Effects of Different Feedback Conditions on Motivational Perceptions, Interest, and Performance. *Journal of Educational Psychology*, 79(4): 474-482.
- Butler, Ruth and Mordecai Nisan. 1986. Effects of No Feedback, Task-Related Comments, and Grades on Intrinsic Motivation and Performance. *Journal of Educational Psychology*, 78(3): 210-216.
- Callahan, Judith Scully, Amy L. Brownlee, Mary D. Brtek and Henry L. Tosi. 2003. Examining the unique effects of multiple motivational sources on task performance. *Journal of Applied Social Psychology*, 33: 2515-2535.
- Chan, Joanne C. Y. and Shui-fong Lam. 2010. Effects of different evaluative feedback on students' self-efficacy in learning. *Instructional Science*, 38(1): 37-58.
- Cherry, Todd L. and Larry V. Ellis. 2005. Does Rank-Order Grading Improve Students Performance? Evidence from a Classroom Experiment. *International Review of Economics Education*, 4(1): 9-19.
- Cohen-Charash, Yoshi and Paul E. Spector. 2001. The Role of Justice in Organizations: A Meta-Analysis. *Organizational Behavior and Human Decision Processes*, 86(2), 278-321.
- Colquitt, Jason A. 2001. On the Dimensionality of Organizational Justice: A Construct Validation of a Measure. *Journal of Applied Psychology*, 86(3): 386-400.
- Covic, Tanya and Mairwen K. Jones. 2008. Is the essay resubmission option a formative or a summative assessment and does it matter as the grades improve? *Assessment & Evaluation in Higher Education*, 33(1): 75-85.
- Deci, Edward L., James P. Connell and Richard M. Ryan. 1989. Self-Determination in a Work Organization. *Journal of Applied Psychology*, 74(4): 580-590.

- Deci, Edward L., Richard M. Ryan and Richard Koestner. 1999. A Meta-Analytical Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. *Psychological Bulletin*, 125(6): 627-668.
- Deci, Edward L. and Richard M. Ryan. 2000. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist*, 55(1): 68-78.
- Duffy, Kirt E. and Robert E. Webber. 1974. On "Relative" Rating Systems. *Personnel Psychology*, 27: 307-311.
- Dysvik, Anders and Bård Kuvaas. 2008. The relationship between perceived training opportunities, work motivation, and employee outcomes. *International Journal of Training and Development*, 12(3): 138-157.
- Early, Christopher P., Gregory B. Northcraft, Cynthia Lee and Terry R. Lituchy. 1990. Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal*, 33(1): 87-105.
- Farndale, Elaine, Veronica Hope-Hailey and Clare Kelliher. 2011. *High commitment performance management: the roles of justice and trust*. *Personnel Review*, 40(1): 5-23.
- Gagné, Marylène and Edward L. Deci. 2005. Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26(4): 331-362.
- Garcia, Stephen M., and Avshalom Tor. 2007. *Rankings, standards, and competition: Task vs. scale comparisons*. *Organizational Behavior and Human Decision Processes*, 102 (1): 95-108.
- Goodman, Jodi S., Robert E. Wood and Margaretha Hendrickx. 2004. Feedback Specificity, Exploration, and Learning. *Journal of Applied Psychology*, 89(2): 248-262.
- Guralnik, Orna, Eyal Rozmarin and Anne So. 2004. Forced Distribution: Is It Right For You? *Human Resource Development Quarterly*, 15(3): 339- 345.
- Hattie, John and Helen Timperley. 2007. The Power of Feedback. *Review of Educational Research*, 77(1): 81-112.
- Hazels Beth and Craig M. Sasse. 2008. Forced Ranking: A review. *SAM Advanced Management Journal*, 73(2): 35-39.
- Ilgen, Daniel R., Cynthia D. Fisher and M. Susan Taylor. 1979. Consequences of Individual Feedback on Behavior in Organizations. *Journal of Applied Psychology*, 64(4): 349-371.
- Kim, Jay S. and W. Clay Hamner. 1976. Effect of performance feedback and goal setting on productivity and satisfaction in an organizational setting. *Journal of Applied Psychology*, 61(1): 48-57.
- Kluger, Avraham N. and Angelo DeNisi. 1996. The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, 119(2): 254-284.
- Kluger, Avraham N. and Angelo DeNisi. 1998. Feedback Interventions: Toward the Understanding of a Double-Edged Sword. *Current Directions in Psychological Science*, 7(3): 67-72.
- Kuvaas, Bård. 2003. Employee ownership and affective organizational commitment: employees' perceptions of fairness and their preference for

- company shares over cash. *Scandinavian Journal of Management*, 19: 193-212.
- Kuvaas, Bård. 2006. Work performance, affective commitment, and work performance: the roles of pay administration and pay level. *Journal of Organizational Behavior*, 27: 365-385.
- Kuvaas, Bård. 2007. Different relationships between perceptions of developmental performance appraisal and work performance. *Personnel Review*, 36(3): 378-397.
- Kuvaas, Bård and Anders Dysvik. 2009. Perceived investment in employee development, intrinsic motivation and work performance. *Human Resource Management Journal*, 19(3): 217-236.
- Larson Jr., James R. 1989. The Dynamic Interplay Between Employees' Feedback-Seeking Strategies and Supervisors' Delivery of Performance Feedback. *Academy of Management Review*, 14(3): 408-422.
- Lawler, Edward E. III. 2002. The Folly of Forced Ranking. *Strategy + Business*, 28: 28-32.
- Lind, Allan, Ruth Kanfer and P. Christopher Early. 1990. Voice, Control and Procedural Justice: Instrumental and Noninstrumental Concerns in Fairness Judgments. *Journal of Personality and Social Psychology*, 59(5): 952-959
- Locke, Edwin A. and Gary P. Latham. 2002. Building a Practically Useful Theory of Goal Setting and Task Motivation. *American Psychologist*, 57(9): 705-717.
- Meisler, Andy. 2003. Dead Man's Curve. *Workforce*, 82: 44-49.
- Meyer, Herbert H. 1975. The Pay for Performance Dilemma. *Organizational Dynamics*, 3: 39-50.
- Olson C.A., and Davis G.M. 2003. Pros and cons of forced ranking and other relative performance rating systems. *Society for Human Resource Management Legal Report*, March.
- Pfeffer, Jeffrey and Robert I. Sutton. 2006. Evidence-Based Management. *Harvard Business Review*, 84(1), 2006, 63-74.
- Piccolo, Ronald F. and Jason A. Colquitt. 2006. Transformational Leadership and Job Behaviors: The Mediating Role of Core Job Characteristics. *Academy of Management Journal*, 49(2): 327-340.
- Podsakoff, Philip M. and Jiing- Lih Farh. 1989. Effects of Feedback Sign and Credibility on Goal Setting and Task Performance. *Organizational Behavior and Human Decision Processes*, 44: 45-67.
- Roch, Sylvia, Angela M. Sternburgh and Pat M. Caputo. 2007. Absolute vs Relative Performance Rating Formats: Implications for fairness and organizational justice. *International Journal of Selection and Assessment*, 15(3): 302-316.
- Ryan, Richard M. 1982. Control and Information in the Intrapersonal Sphere: An Extension of Cognitive Evaluation Theory. *Journal of Personality and Social Psychology*, 43(3): 450-461.
- Salas, Eduardo, Jessica L. Wildman, and Ronald F. Piccolo. 2009. Using Simulation-Based Training to Enhance Management Education. *Academy of Management Learning & Education*, 8 (4): 559-573.

- Schleicher, Deidra J., Rebecca A. Bull and Stephen G. Green. 2009. Rater Reactions to Forced Distribution Rating Systems, *Journal of Management*, 35(4): 899-927.
- Scullen, Steven E., Paul K. Bergey and Lynda Aiman-Smith. 2005. Forced Distribution Rating Systems and the Improvement of Workforce Potential: A Baseline Simulation. *Personnel Psychology*, (58): 1-32.
- Sears David, and Don McDermott. 2003. The Rise and Fall of Rank and Yank. *Information Strategy: The Executive's Journal*, 19: 6-11.
- Simons, Tony and Roberson Quinetta. 2003. Why Managers Should Care About Fairness: The Effects of Aggregate Justice Perceptions on Organizational Outcomes. *Journal of Applied Psychology*, 88(3): 432-443.
- Taras, Maddalena. 2005. Assessment – Summative and Formative – Some Theoretical Reflections. *British Journal of Educational Studies*, 53(4): 466-478.

### **Books**

- Deci, Edwin L. and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.
- Grant, Robert M. 2009. *Contemporary Strategy Analysis*. 7<sup>th</sup> ed. Hoboken, N.J.: John Wiley and Sons Ltd.
- Grote, Dick. 2005. *Forced Ranking: Making performance management work*. Boston, MA: Harvard Business School Press.